DELINEANDO A PESQUISA CLÍNICA

4ª Edição

STEPHEN B. HULLEY

STEVEN R. CUMMINGS

WARREN S. BROWNER

DEBORAH G. GRADY

THOMAS B. NEWMAN

AVISO

Todo esforço foi feito para garantir a qualidade editorial desta obra, agora em versão digital. Destacamos, contudo, que diferenças na apresentação do conteúdo podem ocorrer em função das características técnicas específicas de cada dispositivo de leitura.

importante em relação a medicamentos introduzidos recentemente no mercado farmacêutico ou raramente utilizados.

STEVEN R. CUMMINGS
WARREN S. BROWNER
DEBORAH G. GRADY
THOMAS B. NEWMAN

DELINEANDO A PESQUISA CLÍNICA

4ª edição

Tradução:
Michael Schmidt Duncan
André Garcia Islabão

Revisão técnica:
Michael Schmidt Duncan
Médico de Família e Comunidade. Preceptor do Programa de Residência em Medicina de Família e Comunidade da Secretaria Municipal de Saúde do Rio de Janeiro. Editor da Revista Brasileira de Medicina de Família e Comunidade.

2015

Obra originalmente publicada sob o título *Designing clinical research*, 4th Edition
ISBN 978-1-60831-804-9

Gerente editorial: *Letícia Bispo de Lima*

Colaboraram nesta edição:

Editora: *Mirian Raquel Fachinetto Cunha*

Capa: *Márcio Monticelli*

Preparação de originais: *Janice Ribeiro de Souza e Heloísa Stefan*

Leitura final: *Magda Regina Schwartzhaupt*

Editoração: *Techbooks*

Produção digital: *Know-How Editorial*

D356 Delineando a pesquisa clínica [recurso eletrônico] / Stephen B.
Hulley ... [et al.] ; tradução e revisão técnica : Michael

Schmidt Duncan ; tradução: André Garcia Islabão. – 4. ed. –
Porto Alegre : Artmed, 2015.
e-PUB.

Editado como livro impresso em 2015.
ISBN 978-85-8271-203-0

1. Epidemiologia. I. Hulley, Stephen B. .

CDU 616-036.22

Catalogação na publicação: Poliana Sanchez de Araujo – CRB 10/2094

Av. Jerônimo de Ornelas, 670 – Santana
90040-340 – Porto Alegre – RS
Fone: (51) 3027-7000 Fax: (51) 3027-7070

Unidade São Paulo
Av. Embaixador Macedo Soares, 10.735 – Pavilhão 5 – Cond. Espace Center
Vila Anastácio – 05095-035 – São Paulo – SP
Fone: (11) 3665-1100 Fax: (11) 3667-1333

SAC 0800 703-3444 – **www.grupoa.com.br**

Autores

Stephen B. Hulley, MD, MPH
Professor and Chair, Emeritus
Department of Epidemiology & Biostatistics
School of Medicine, University of California, San Francisco

Steven R. Cummings, MD
Founding Director, San Francisco Coordinating Center
Senior Scientist, California Pacific Medical Center Research Institute
Professor Emeritus, Department of Medicine, and of Epidemiology & Biostatistics
School of Medicine, University of California, San Francisco

Warren S. Browner, MD, MPH
Chief Executive Officer, California Pacific Medical Center
Adjunct Professor, Department of Epidemiology & Biostatistics
School of Medicine, University of California, San Francisco

Deborah G. Grady, MD, MPH
Professor of Medicine
Associate Dean for Clinical and Translational Research
School of Medicine, University of California, San Francisco

Thomas B. Newman, MD, MPH
Professor of Epidemiology & Biostatistics, and of Pediatrics
Chief, Division of Clinical Epidemiology
Attending Physician, Department of Pediatrics
School of Medicine, University of California, San Francisco

Colaboradores

Bernard Lo, MD

President, The Greenwall Foundation
Professor of Medicine, Emeritus
Director of Program in Medical Ethics, Emeritus
University of California, San Francisco

Michael A. Kohn, MD, MPP
Associate Professor of Epidemiology and Biostatistics
School of Medicine, University of California, San Francisco
Attending Physician, Emergency Department
Mills-Peninsula Medical Center, Burlingame, California

Norman Hearst, MD, MPH
Professor of Family and Community Medicine
School of Medicine, University of California, San Francisco
Attending Physician, University of California Medical Center
San Francisco, California

Thomas Edward Novotny, MD, MPH
Professor and Associate Director for Border and Global Health
Graduate School of Public Health
San Diego State University, San Diego, California

Dedicatória

Para nossas famílias e nossos alunos

Agradecimentos

Somos gratos à Fundação Andrew W. Mellon pelo fato de ter reunido os cinco autores deste livro, há 30 anos, para começar a jornada de cinco anos do desenvolvimento dos materiais de ensino que se tornaram nossa 1ª edição; ao nosso editor por nos convencer firmemente a fazer a 4ª edição até que nossa resistência tivesse se esgotado, e por oferecer profissionais excepcionalmente talentosos e apoiadores para nos ajudar a realizar a tarefa; a nossas famílias por seu apoio paciente à medida que trabalhávamos nesta obra; a vários colegas da UCSF e outros, cujas ideias influenciaram as nossas; a nossos alunos ao longo dos anos, cujas conquistas foram bonitas de ver e estimularam nosso pensamento; e a nossos leitores, que colocam este livro em uso.

Sumário

Introdução

Esta 4ª edição de *Delineando a pesquisa clínica* marca o 25º aniversário da publicação da nossa 1ª edição. A presente obra tornou-se o livro-texto mais usado da área, com mais de 130.000 cópias vendidas e edições em outros idiomas, como espanhol, português, árabe, chinês, coreano e japonês. Elaboramos este manual para pesquisa clínica em todas as suas nuances: ensaios clínicos, epidemiologia observacional, ciência translacional, pesquisa orientada ao paciente, ciência comportamental e pesquisa em assistência à saúde. Usamos termos e princípios da epidemiologia, apresentamos informação conceitual avançada de forma prática e acessível ao leitor e sugerimos abordagens de bom senso para muitas decisões envolvidas ao se delinear um estudo.

Muitos de nossos leitores são médicos, enfermeiros, farmacêuticos e outros cientistas da saúde que estão iniciando sua carreira em pesquisa clínica e usam este livro como guia para delinear e conduzir seus estudos. Muitos outros são médicos em programas de residência ou alunos de faculdades da área de saúde – medicina, enfermagem, farmácia e saúde pública, entre outros – que usam a presente obra para tornarem-se leitores exigentes com uma compreensão das forças e limitações dos estudos de pesquisa que informam a prática da clínica baseada em evidências. Um terceiro público consiste em estudantes ainda não graduados que estão se preparando para entrar nessa área e que têm interesse em conhecer o mundo da pesquisa clínica.

O que há de novo nesta edição? A inovação mais visível é a adição de cor, que, além da melhora estética, irá acelerar a compreensão dos conteúdos codificados por seu uso. Com relação ao conteúdo, as revisões importantes desta 4ª edição incluíram a atualização e uma maior concisão dos textos, figuras e tabelas em todos os capítulos; muitos exemplos e referências novos; e novas seções cobrindo avanços recentes da área. Por

exemplo:

- Os capítulos sobre estudos observacionais foram reorganizados, com um novo capítulo inteiro dedicado a vários delineamentos de caso-controle, incluindo a abordagem de incidência-densidade para lidar com mudanças nos níveis de fatores de risco e diferenças no tempo de seguimento.
- Os capítulos sobre ensaios clínicos tiveram um aumento na seção sobre ensaios de não inferioridade, que se tornaram populares em estudos sobre eficácia comparativa, e tratam de análises de subgrupo e modificação do efeito de uma forma mais completa.
- O capítulo sobre estudos de testes médicos apresenta uma nova seção sobre a prática crescente de desenvolver regras de predição clínica.
- O capítulo sobre a utilização de bancos de dados já existentes enfatiza opções interessantes para pesquisadores iniciantes poderem publicar com rapidez e baixo custo.
- O capítulo sobre ética em pesquisa foi atualizado para refletir as políticas atuais sobre sequenciamento do genoma inteiro e outros tópicos, com novos casos que ilustram a resolução de dilemas éticos na pesquisa clínica.
- O capítulo sobre gerenciamento de dados foi extensamente atualizado com as últimas abordagens baseadas na web.
- O capítulo sobre obtenção de financiamento apresenta estratégias relativas aos novos requisitos para concessão de financiamento, bem como atualizações sobre financiamento por fundações e patrocinadores corporativos.

Esta edição vem acompanhada por um website atualizado (em inglês) em www.epibiostat.ucsf.edu/dcr/ que contém materiais para ensinar delineamento de pesquisa clínica, incluindo links para o programa detalhado dos workshops de 4 e 7 semanas que apresentamos a 300 pessoas em treinamento a cada ano na UCSF. Há também notas para os

instrutores dos workshops que serão úteis àqueles que ensinam este conteúdo, além de links para nosso programa de mestrado Training In Clinical Research (TICR) na UCSF, com mais de 30 outros cursos e seus materiais. Além disso, há ferramentas valiosas para investigadores, incluindo uma excelente calculadora interativa de tamanho de amostra.

Muitas coisas *não* mudaram nesta 4ª edição. Este continua sendo um livro simples que deixa de fora tecnicalidades desnecessárias e que convida o investigador a se concentrar nas coisas importantes: como encontrar uma boa questão de pesquisa e como planejar um delineamento eficiente, efetivo e ético. Os capítulos sobre estimativa do tamanho de amostra continuam a desmistificar o processo e capacitam os leitores com treinamento mínimo em estatística para fazer os cálculos sozinhos, de maneira pensada e sem necessidade de se debater com fórmulas. O livro continua funcionando melhor quando combinado com o ingrediente essencial que é ter um ou mais mentores a longo prazo. Ele continua não tratando de áreas importantes sobre como analisar, apresentar e publicar os achados de pesquisa clínica – tópicos que nossos leitores podem buscar em outros livros (p. ex., 1– 4).

O processo de tornar-se um cientista clínico independente pode ser desafiador, especialmente no que se refere a vencer o obstáculo de obter um financiamento substancial pela primeira vez. Mas é gratificante que muitos de nossos primeiros alunos que usaram este livro conseguiram atingir esse objetivo, descobriram que gostam de fazer pesquisa e se estabeleceram em uma grande carreira. Para aqueles com mentes inquisitivas, a busca pela verdade pode tornar-se uma fascinação para a vida inteira. Para perfeccionistas e artesãos, há desafios intermináveis para criar estudos elegantes que respondem a questões de forma conclusiva, pequenos e grandes, razoáveis em termos de tempo e custo financeiro. Os investigadores que gostam de trabalhar em equipe desenvolverão relacionamentos recompensadores com colegas, funcionários e estudantes, bem como amizades com colaboradores trabalhando no mesmo campo em lugares distantes. E para aqueles com ambição de fazer uma contribuição duradoura para a sociedade, há a perspectiva de que com habilidade e tenacidade eles participarão dos avanços crescentes na prática da saúde clínica e pública que é a ordem

natural da nossa ciência.

REFERÊNCIAS

1. Vittinghoff E, Glidden DV, Shiboski SC, et al. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*, 2nd ed. New York: Springer-Verlag, 2011.
2. Katz MH. *Multivariable analysis: a practical guide for clinicians and public health researchers*, 3rd ed. New York: Cambridge University Press, 2011.
3. Newman TB, Kohn MA. *Evidence-based diagnosis*. Cambridge, MA: Cambridge University Press, 2009.
4. Browner WS. *Publishing and presenting clinical research*, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2012.

SEÇÃO I

Ingredientes básicos

CAPÍTULO 1

Introdução: anatomia e fisiologia da pesquisa clínica

Stephen B. Hulley, Thomas B. Newman e Steven R. Cummings

Este capítulo introduz a pesquisa clínica sob duas perspectivas distintas, estabelecendo linhas temáticas que andam juntas no decorrer do livro. A primeira é a **anatomia** da pesquisa – de que ela é feita. Isso inclui os elementos tangíveis do plano de estudo, como questão de pesquisa, delineamento (desenho), sujeitos, medidas, cálculo do tamanho de amostra e assim por diante. A meta do investigador é montar esses componentes de forma que o projeto se torne factível e eficiente.

A outra linha temática é a **fisiologia** da pesquisa – como ela funciona. Os estudos são úteis na medida em que possibilitam inferências válidas, primeiro sobre o que ocorreu na amostra do estudo e então sobre a generalização desses eventos para indivíduos externos ao estudo. Assim, a meta é minimizar erros, aleatórios ou sistemáticos, que ameacem as conclusões advindas dessas inferências.

A separação desses dois temas é artificial, assim como a anatomia do corpo humano não faz muito sentido sem uma compreensão de sua fisiologia. Mas a separação traz a mesma vantagem: torna mais claro o

nosso entendimento sobre um tema complexo.

ANATOMIA DA PESQUISA: DE QUE ELA É FEITA

A estrutura de um projeto de pesquisa é descrita em seu **protocolo**, o plano escrito do estudo. Os protocolos são instrumentos usados na solicitação de recursos financeiros e na avaliação pelo Comitê de Ética em Pesquisa (CEP), mas também têm outra função científica vital: ajudam o investigador a organizar sua pesquisa de forma lógica, objetiva e eficiente. A Tabela 1.1 apresenta os elementos que compõem um protocolo. Neste capítulo, introduziremos o conjunto desses elementos; nos seguintes, desenvolveremos cada um deles; por fim, no Capítulo 19, juntaremos todas as peças, de forma integrada.

TABELA 1.1 Anatomia da pesquisa: o plano de estudo

ELEMENTO	OBJETIVO
Questões de pesquisa	Quais questões o estudo abordará?
Relevância (background)	Por que essas questões são importantes?
Delineamento Eixo temporal Abordagem epidemiológica	Como o estudo é estruturado?
Sujeitos Critérios de seleção Desenho amostral	Quem são os sujeitos e como eles serão selecionados?
Variáveis Variáveis preditoras Variáveis confundidoras Variáveis de desfecho	Quais medições serão realizadas?
Aspectos estatísticos Hipóteses Tamanho de amostra Abordagem analítica	Qual é o tamanho do estudo e como ele será analisado?

Questão de pesquisa

A **questão de pesquisa** é o objetivo do estudo, a incerteza que o investigador deseja resolver. As questões de pesquisa partem de uma preocupação geral, que necessita ser reduzida a um tópico concreto e factível de ser estudado. Por exemplo, observe a questão de pesquisa a seguir:

- *As pessoas deveriam comer mais peixe?*

Este é um bom ponto de partida, porém a questão deve ser objetivada antes que se possa começar a planejar o estudo. Isso significa separar as partes que constituem a questão e escolher uma ou duas delas para, então, elaborar o protocolo:

- *Com que frequência os norte-americanos comem peixe?*
- *Comer mais peixe diminui o risco de desenvolver doença cardiovascular?*
- *Existe um risco aumentado de intoxicação por mercúrio quando pessoas idosas passam a comer mais peixe?*
- *Os suplementos de óleo de peixe têm os mesmos efeitos sobre doença cardiovascular que o peixe presente na dieta?*
- *Que suplementos de óleo de peixe não fazem o seu hálito cheirar a peixe?*

Uma boa questão de pesquisa deve passar no teste do "E daí?". A resposta à questão deve contribuir para nosso estado de conhecimento. O acrônimo **FINER** reúne as cinco características básicas de uma boa questão de pesquisa: ela deve ser **factível**, **interessante**, **nova** (inovadora, original), **ética** e **relevante** (Capítulo 2).

Relevância

Uma breve seção sobre **relevância** no protocolo mostra como o estudo se insere em um contexto maior e apresenta a sua justificativa (*rationale*). O que se sabe sobre o tema? Por que a questão de pesquisa é importante? Que respostas o estudo fornecerá? Essa seção apresenta pesquisas anteriores relevantes (incluindo o trabalho anterior desenvolvido pelo próprio pesquisador), indicando seus problemas e questões pendentes. Esclarece também como os achados poderão ajudar a resolver essas incertezas, estabelecer uma nova compreensão científica ou influenciar decisões ou diretrizes clínicas e de saúde pública. Com frequência, a revisão da literatura e a síntese feita para a seção sobre relevância levam o

pesquisador a modificar a questão de pesquisa original.

Delineamento (desenho)

O **delineamento** de um estudo é um tópico complexo. Antes de tudo, é fundamental que se escolha entre desempenhar um papel passivo, simplesmente fazendo aferições nos sujeitos do estudo – um **estudo observacional** –, ou aplicar uma intervenção e examinar seus efeitos – um **ensaio clínico** (Tabela 1.2). Entre os estudos observacionais, os dois delineamentos mais comuns são o **estudo de coorte**, em que um grupo de sujeitos é seguido ao longo do tempo, e o **estudo transversal**, em que as observações são feitas em uma única ocasião. Os estudos de coorte podem ainda ser divididos em **prospectivos**, que iniciam no presente e seguem os sujeitos no tempo, e **retrospectivos**, que examinam dados coletados ao longo de um período de tempo no passado. Uma terceira opção comum é o estudo do tipo **caso-controle**, no qual o investigador compara um grupo de sujeitos que têm uma doença ou outro desfecho com outro grupo de sujeitos que não a têm. Entre as opções de ensaio clínico, o **ensaio clínico randomizado cego** é em geral o melhor delineamento, mas ensaios clínicos não cegos ou não randomizados podem ser a única opção factível para determinadas questões de pesquisa.

TABELA 1.2 Exemplos de delineamentos comuns na pesquisa clínica que podem ser usados no estudo que avalia se o consumo de peixe reduz o risco de doença coronariana

DELINEAMENTO	CARACTERÍSTICA PRINCIPAL	EXEMPLO
	Estudos observacionais	
Estudo de coorte	Um grupo de pessoas identificadas no início e seguidas ao longo do tempo	O investigador mede o consumo de peixe na linha de base e examina periodicamente os sujeitos nas visitas de seguimento para avaliar se aqueles que consomem mais peixe têm menor número de eventos coronarianos.
Estudo transversal	Um grupo examinado em um determinado momento no tempo	O investigador entrevista um grupo de sujeitos a respeito da história atual ou prévia de consumo de peixe, correlacionando os resultados com a história de doença coronariana e com o escore atual de cálcio nas coronárias.
Estudo de caso-controle	Dois grupos selecionados a partir da presença ou ausência de um desfecho	O investigador examina um grupo de pacientes com doença coronariana (os "casos") e os compara com um grupo sem doença coronariana (os "controles"),

		perguntando sobre o consumo de peixe no passado.
	Ensaio clínico	
Ensaio clínico randomizado cego	Dois grupos gerados por um processo aleatório e uma intervenção mascarada (cegada)	O investigador aloca sujeitos aleatoriamente para receber suplementos de óleo de peixe ou um placebo idêntico em aparência e então acompanha os dois grupos por vários anos para observar a incidência de doença coronariana.

Nenhuma abordagem é sempre melhor do que as demais, e cada questão de pesquisa requer uma escolha racional sobre que desenho seria o mais eficiente na obtenção de uma resposta adequada. O ensaio clínico randomizado cego é considerado o padrão-ouro para estabelecer causalidade e efetividade de intervenções, mas, em muitos casos, um estudo observacional é a melhor opção ou a única opção factível. O custo relativamente baixo de estudos de caso-controle e a sua aplicabilidade para desfechos incomuns os tornam atrativos para muitas questões de pesquisa. Considerações especiais aplicam-se à escolha de delineamentos para testes diagnósticos. Toda essa problemática será discutida nos Capítulos 7 a 12, cada um deles abordando um determinado conjunto de delineamentos.

Uma sequência típica na investigação de uma questão de pesquisa inicia com estudos observacionais do tipo **descritivo**. Esses estudos exploram a topografia do terreno, por exemplo, as distribuições das características de saúde e das doenças em uma população.

- *Em média, quantas porções de peixe por semana estão presentes na dieta de norte-americanos com história de doença coronariana?*

Os estudos descritivos são normalmente seguidos ou acompanhados por estudos **analíticos**, que avaliam as associações para realizar inferências sobre relações de causa-efeito.

- *As pessoas com doença coronariana que comem muito peixe têm menor risco de infarto do miocárdio recorrente do que pessoas com doença coronariana que raramente comem peixe?*

O último passo é, em geral, um **ensaio clínico** para estabelecer os efeitos de uma intervenção.

- *O tratamento com cápsulas de óleo de peixe reduz a mortalidade geral em indivíduos com doença coronariana?*

Os ensaios clínicos geralmente ocorrem em uma etapa mais avançada na sequência da investigação, pois tendem a ser mais difíceis e caros, além de responderem de forma mais definitiva a questões mais específicas que costumam surgir dos achados de estudos observacionais.

É aconselhável caracterizar o estudo em uma única frase que resuma o delineamento e a questão de pesquisa. Se o estudo apresentar duas fases importantes, o delineamento de cada uma delas deverá ser mencionado, como a seguir:

- *Trata-se de um estudo transversal sobre hábitos alimentares em indivíduos de 50 a 69 anos com história de doença coronariana, seguido de um estudo de coorte prospectivo sobre a associação entre consumo de peixe e um menor risco de eventos isquêmicos subsequentes.*

Essa frase é o análogo na pesquisa da frase inicial de um relato de um residente de medicina sobre uma paciente recentemente hospitalizada: “Essa policial branca de 62 anos estava em bom estado de saúde até duas horas antes de vir ao hospital, quando desenvolveu ‘dor em aperto’ no peito com irradiação para o ombro esquerdo”.

Alguns delineamentos, no entanto, não se encaixam facilmente nas categorias listadas e classificá-los usando uma única frase pode ser um exercício surpreendentemente difícil. Mesmo assim, vale o esforço. Uma descrição concisa do tipo de estudo torna as ideias do investigador mais claras e ajuda a orientar os colaboradores e consultores.

Sujeitos do estudo

Duas decisões importantes devem ser tomadas na escolha dos sujeitos do estudo (Capítulo 3). A primeira é especificar os **critérios de inclusão** e de **exclusão** que definem a população-alvo: os *tipos* de pessoas mais adequados à questão de pesquisa. A segunda é definir como **recrutar** um *número* apropriado de pessoas de um subconjunto acessível dessa população, as quais serão os reais sujeitos do estudo. Por exemplo, o estudo sobre consumo de peixe em pessoas com doença coronariana pode

identificar indivíduos vistos em um ambulatório com diagnóstico de infarto do miocárdio, angioplastia ou cirurgia de revascularização miocárdica registrado no seu prontuário eletrônico. As decisões sobre que pacientes estudar devem levar em conta as vantagens e desvantagens das diferentes opções; estudar uma amostra aleatória de todas as pessoas com doença coronariana nos Estados Unidos (ou pelo menos de vários diferentes Estados e cenários de cuidados médicos) poderia aumentar a capacidade de **generalização**, mas seria extremamente dispendioso e difícil.

Variáveis

Outro conjunto importante de decisões que devem ser tomadas no delineamento de qualquer estudo se refere à escolha de que variáveis medir (Capítulo 4). Por exemplo, um estudo sobre o peixe presente na dieta poderia perguntar sobre o consumo de diferentes tipos de peixe contendo diferentes níveis de ácidos graxos V-3, bem como incluir questões sobre aspectos como o tamanho das porções, se o peixe foi servido frito ou assado e o uso de suplementos de óleo de peixe.

Em um estudo analítico, o investigador estuda as associações entre duas ou mais variáveis para predizer os desfechos e fazer inferências sobre causa e efeito. Ao considerar a associação entre duas variáveis, a que precede (ou é pressuposta biologicamente como antecedente) é denominada **variável preditora**; a outra, **variável de desfecho**.[1] A maioria dos estudos observacionais tem muitas variáveis preditoras (p. ex., *idade, raça, sexo, história de tabagismo, consumo de peixe e de suplementos de óleo de peixe*) e muitas de desfecho (p. ex., *infartos, acidentes vasculares encefálicos [AVEs], qualidade de vida, odor desagradável*).

Os ensaios clínicos estudam os efeitos de uma **intervenção** (tipo especial de variável preditora manipulada pelo investigador), como, por exemplo, o *tratamento com cápsulas de óleo de peixe*. Esse delineamento permite observar os efeitos dessa intervenção na variável de desfecho, usando a **randomização** para minimizar a influência de **variáveis confundidoras** – outros preditores do desfecho, como o *tabagismo* ou a *renda familiar*, que poderiam estar associados ao consumo de peixe e assim confundir a interpretação dos achados.

Aspectos estatísticos

O investigador deve planejar como estimar o tamanho da amostra e como manejar e analisar os dados. Isso geralmente envolve a especificação de uma **hipótese** (Capítulo 5):

Hipótese: *Mulheres de 50 a 69 anos com doença coronariana que tomam suplementos à base de óleo de peixe têm menor risco de infarto do miocárdio recorrente do que aquelas que não os tomam.*

Esta é uma versão da questão de pesquisa que fornece as bases para o teste de **significância estatística**. A especificação da hipótese também permite que o investigador estime o **tamanho da amostra** – o número de sujeitos necessários para observar a diferença esperada no desfecho entre os grupos, com uma probabilidade razoável (um atributo conhecido como **poder estatístico**) (Capítulo 6). Estudos inteiramente descritivos (p. ex., *que proporção dos indivíduos com doença coronariana tomam suplementos de óleo de peixe?*) não envolvem testes de significância estatística e portanto não requerem a formulação de uma hipótese; nesses casos, a abordagem análoga seria estimar o número de sujeitos necessário para alcançar um nível aceitável de precisão nos cálculos de **intervalos de confiança** para médias, proporções ou outras estatísticas descritivas.

FISIOLOGIA DA PESQUISA: COMO ELA FUNCIONA

A meta da pesquisa clínica é **inferir** a partir dos resultados do estudo sobre a natureza da verdade no universo. Dois tipos principais de inferências estão envolvidos na interpretação de um estudo (conforme ilustrado na porção superior da Figura 1.1, da direita para a esquerda). A primeira inferência refere-se à **validade interna**, ou seja, até que ponto estão corretas as conclusões do investigador sobre o que realmente ocorreu no estudo. A segunda refere-se à **validade externa** (também denominada **capacidade de generalização**), ou seja, o quanto essas conclusões se aplicam a pessoas e eventos externos ao estudo.

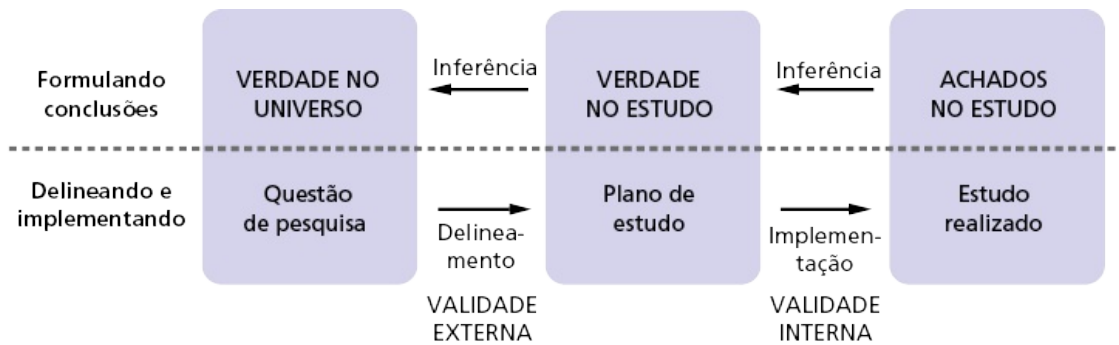



FIGURA 1.1 O processo de delinear e implementar um projeto de pesquisa prepara o terreno para se chegar a conclusões a partir dos achados.

Ao planejar um estudo, o investigador inverte essa sequência, partindo da esquerda para a direita na porção inferior da Figura 1.1, com o objetivo de maximizar a validade das inferências que poderão ser feitas ao final do estudo. Para tanto, **desenha um plano de estudo** no qual a escolha da questão de pesquisa, dos sujeitos e das aferições é feita de modo a aumentar a validade externa do estudo e a conduzir a uma **implementação** que assegure um alto grau de validade interna. Na seções a seguir, abordaremos primeiramente os tópicos de delineamento e implementação, então, abordaremos os erros que ameaçam a validade das inferências da pesquisa clínica.

Delineando o estudo

Considere a questão de pesquisa descritiva a seguir:

Qual é a prevalência de consumo diário de suplementos de óleo de peixe em indivíduos com doença coronariana?

Essa questão não pode ser respondida com exatidão, porque seria impossível estudar todos os pacientes com doença coronariana e porque os meios para verificar se uma pessoa tem doença coronariana e está tomando suplementos de óleo de peixe são imperfeitos. Portanto, o investigador deve formular uma questão semelhante que possa ser respondida pelo estudo proposto, tal como:

Em uma amostra de pacientes vistos no ambulatório do investigador com diagnóstico prévio de doença coronariana e que respondem a um questionário enviado pelo correio, que proporção relata estar tomando suplementos de óleo de peixe diariamente?

A Figura 1.2 ilustra a transformação de uma questão de pesquisa em um plano de estudo. Um componente importante dessa transformação é a escolha de uma **amostra** de sujeitos que represente a **população**. O grupo de sujeitos especificado no protocolo só pode ser uma amostra da população de interesse, porque barreiras de ordem prática limitam o estudo da população inteira. A decisão de estudar pacientes vistos no ambulatório do investigador admite essa limitação. Essa é uma amostra factível de ser estudada, mas tem a desvantagem de que pode produzir uma prevalência de consumo de óleo de peixe diferente da encontrada no universo de pacientes com doença coronariana.

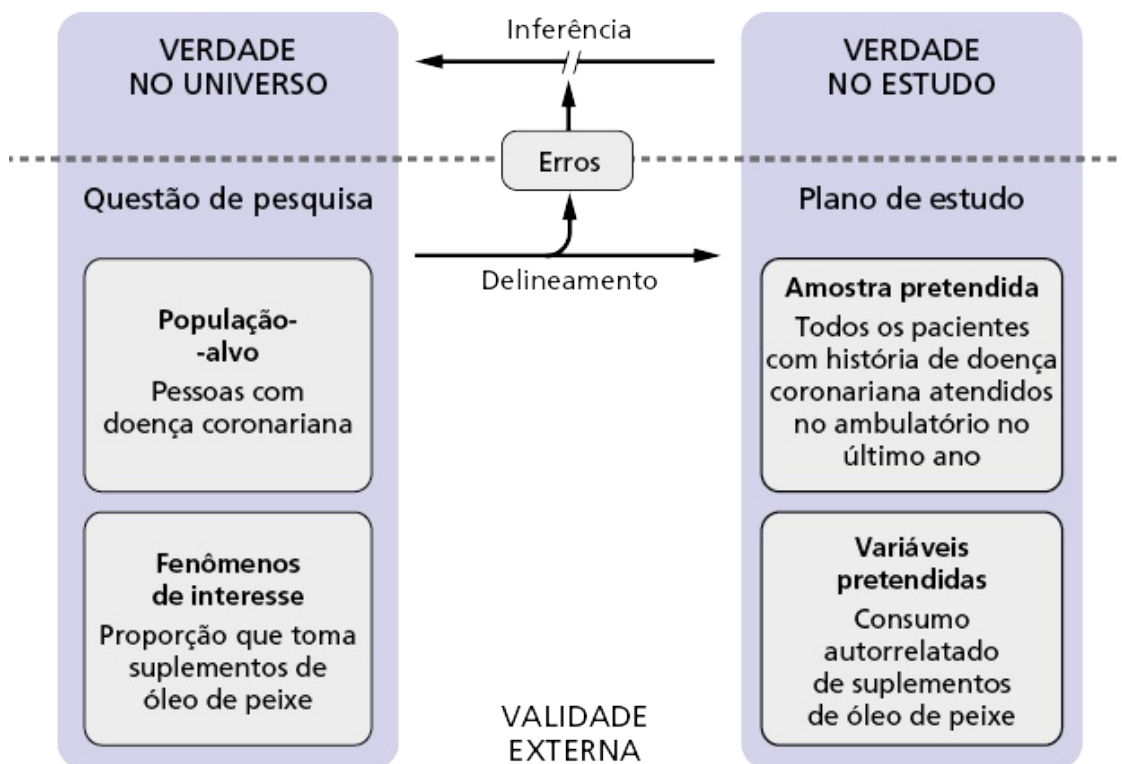



FIGURA 1.2 Erros de delineamento e validade externa: Quando a amostra e as variáveis pretendidas não representam suficientemente a população-alvo e os fenômenos de interesse, esses erros poderão distorcer as inferências sobre o que realmente ocorreu na população.

O outro componente importante da transformação é a escolha das **variáveis** que irão representar os **fenômenos de interesse**. As variáveis especificadas no plano de estudo são geralmente substitutas (*proxies*) para esses fenômenos. A decisão de usar um questionário para avaliar o

consumo de óleo de peixe é uma forma rápida e de baixo custo para coletar informações, mas é provável que não seja inteiramente exata, uma vez que as pessoas geralmente não lembram com exatidão ou não registram de forma adequada o seu consumo médio semanal de óleo de peixe.

Em suma, cada uma das diferenças na Figura 1.2 entre a questão de pesquisa e o plano de estudo tem o objetivo de tornar o estudo mais prático. No entanto, o custo do ganho em praticidade é o risco de as escolhas feitas no delineamento levarem o estudo a produzir uma conclusão incorreta ou enganadora, porque ele é desenhado para responder a uma questão que difere da questão de pesquisa de interesse.

Implementando o estudo

Voltando à Figura 1.1, o lado direito refere-se à **implementação** e ao grau em que a pesquisa, na forma realizada, reflete o plano de estudo. Em outras palavras, trata-se do problema de responder erroneamente à questão de pesquisa porque a forma de selecionar a amostra ou realizar as aferições diferiu da forma que havia sido delineada (Figura 1.3).

A amostra de sujeitos selecionada para o estudo quase sempre difere da amostra pretendida. Por exemplo, o plano de estudar todos os pacientes elegíveis com doença coronariana em um ambulatório poderia ser comprometido pelo registro incompleto dos diagnósticos no prontuário eletrônico, pelo envio de questionários a endereços errados e pela recusa em participar. Os indivíduos que foram identificados e concordaram em participar podem apresentar uma prevalência diferente de consumo de óleo de peixe do que aqueles que não foram identificados ou se recusaram a participar. Além desses problemas com os sujeitos, as próprias aferições podem diferir das aferições pretendidas. Por exemplo, se a formatação do questionário for pouco clara, os pacientes podem ficar confusos e marcar a resposta errada, ou podem simplesmente omitir a questão por engano.

Essas diferenças entre o plano de estudo e o estudo realizado podem alterar a resposta para a questão de pesquisa. A Figura 1.3 ilustra o fato importante de que erros na implementação do estudo são outras razões comuns (além dos erros de delineamento) de se chegar à resposta errada à questão de pesquisa.

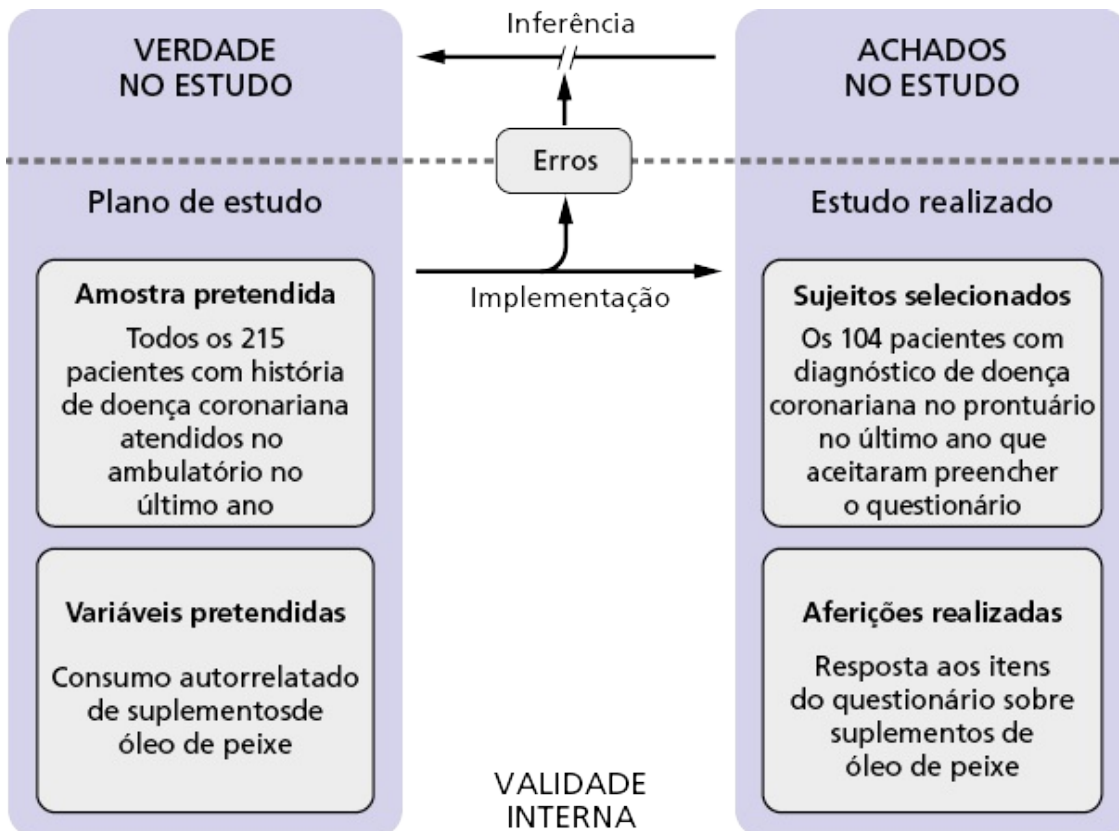



FIGURA 1.3 Erros de implementação e validade interna: Se os sujeitos selecionados e as aferições realizadas não representarem de forma suficiente a amostra e as variáveis pretendidas, esses erros poderão distorcer as inferências sobre o que ocorreu no estudo.

Inferências causais

Um tipo especial de problema de validade surge em estudos que examinam a **associação** entre uma variável preditora e uma variável de desfecho para produzir inferências causais. Se um estudo de coorte encontra uma associação entre o consumo de peixe e eventos coronarianos, isso representa causa e efeito ou o consumo de peixe é um mero observador inocente no meio de uma teia de causalidade que envolve outras variáveis? Reduzir a probabilidade de fatores de **confusão** e outras explicações rivais é um dos maiores desafios no delineamento de um estudo observacional (Capítulo 9).

Os erros da pesquisa

Reconhecendo que nenhum estudo é inteiramente livre de erros, a meta é maximizar a validade das inferências causais sobre o que foi observado na

amostra de estudo para concluir sobre o que está acontecendo na população.

Inferências errôneas podem ser trabalhadas na etapa de análise da pesquisa, mas as melhores estratégias centram-se no delineamento e na implementação do estudo (Figura 1.4), pois, desde que práticas, previnem a ocorrência de erros já no início.

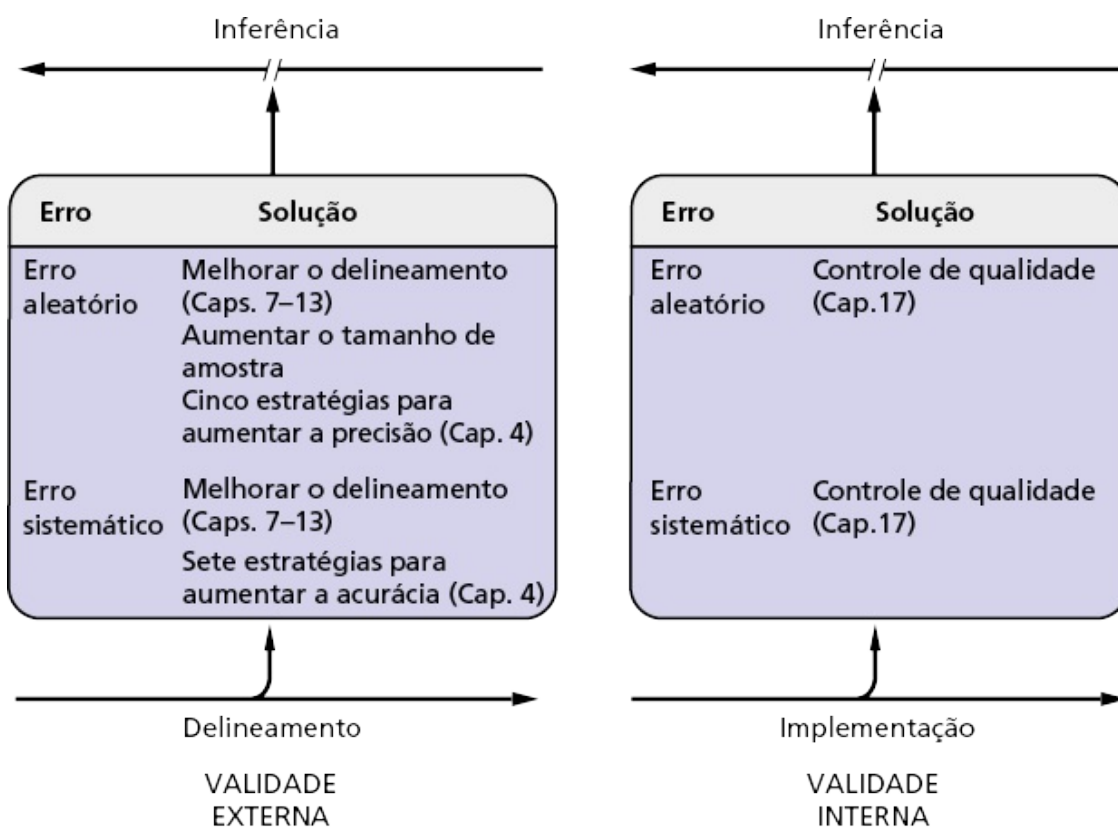



FIGURA 1.4 Erros da pesquisa. Essa versão expandida dos quadros referentes a erros nas Figuras 1.2 e 1.3 revela estratégias para controlar o erro aleatório e erro sistemático nas fases de delineamento e implementação do estudo.

Os dois principais tipos de erros que interferem nas inferências da pesquisa são o erro aleatório e o erro sistemático. Essa distinção é importante, pois as estratégias para minimizá-los são bastante diferentes.

Erro aleatório é um resultado errado devido ao **acaso**. É uma fonte de variação com igual probabilidade de distorcer as aferições em ambas as direções. Se a prevalência real de consumo diário de suplementos de óleo de peixe nas centenas de pacientes com doença coronariana de 50 a 69 anos atendidos no ambulatório do investigador é de 20%, uma amostra

bem-delineada de 100 pacientes dessa população poderia conter exatamente 20 pacientes que tomam esses suplementos. No entanto, é mais provável que a amostra contenha um número próximo, como 18, 19, 21 ou 22. Eventualmente, o acaso poderia produzir um número substancialmente diferente, como 12 ou 28. Entre as várias técnicas para reduzir a influência do erro aleatório (Capítulo 4), a mais simples é a de aumentar o tamanho da amostra. O uso de uma amostra maior diminui a probabilidade de um resultado substancialmente errado, aumentando a **precisão** da estimativa, ou seja, o grau em que a prevalência observada se aproxima de 20% toda vez em que uma amostra é sorteada.

Erro sistemático é um resultado errado devido a um **viés** (fonte de variação que distorce os achados do estudo para uma determinada direção). Um exemplo seria a decisão, na Figura 1.3, de usar pacientes que frequentam o ambulatório do investigador, onde o tratamento tende a refletir o interesse do pesquisador nesse assunto e onde os médicos estariam mais propensos a recomendar o consumo de óleo de peixe do que a média dos médicos. Aumentar o tamanho da amostra não reduz o erro sistemático. A melhor forma de melhorar a **acurácia** da estimativa (grau em que se aproxima do valor real) é delinear o estudo de modo a reduzir a magnitude dos vários vieses. Uma alternativa é buscar informações adicionais para avaliar a importância de possíveis vieses. Um exemplo poderia ser comparar os resultados com aqueles de uma segunda amostra de pacientes com doença coronariana selecionados de um outro contexto; por exemplo, avaliar se os achados em pacientes atendidos em um ambulatório de cardiologia diferem daqueles de pacientes vistos em um ambulatório de atenção primária.

Os exemplos de erros aleatórios e sistemáticos nos dois parágrafos anteriores são componentes do **erro amostral**, que ameaça inferências sobre os sujeitos do estudo para a população. Os erros aleatórios e sistemáticos podem também contribuir para **erros de aferição**, ameaçando as inferências das aferições do estudo para os fenômenos de interesse. Um exemplo de erro de aferição aleatório é a variação na resposta a um questionário administrado a um paciente em diferentes ocasiões. Um exemplo de erro de aferição sistemático é a subestimativa da prevalência de consumo de óleo de peixe por falta de clareza na formulação da pergunta. Outras estratégias para controlar essas fontes de

erro são apresentadas nos Capítulos 3 e 4.

Os conceitos apresentados anteriormente são resumidos na Figura 1.5. Em suma, obter uma resposta correta à questão de pesquisa significa delinear e implementar o estudo de tal forma que minimize a magnitude do erro inferencial.

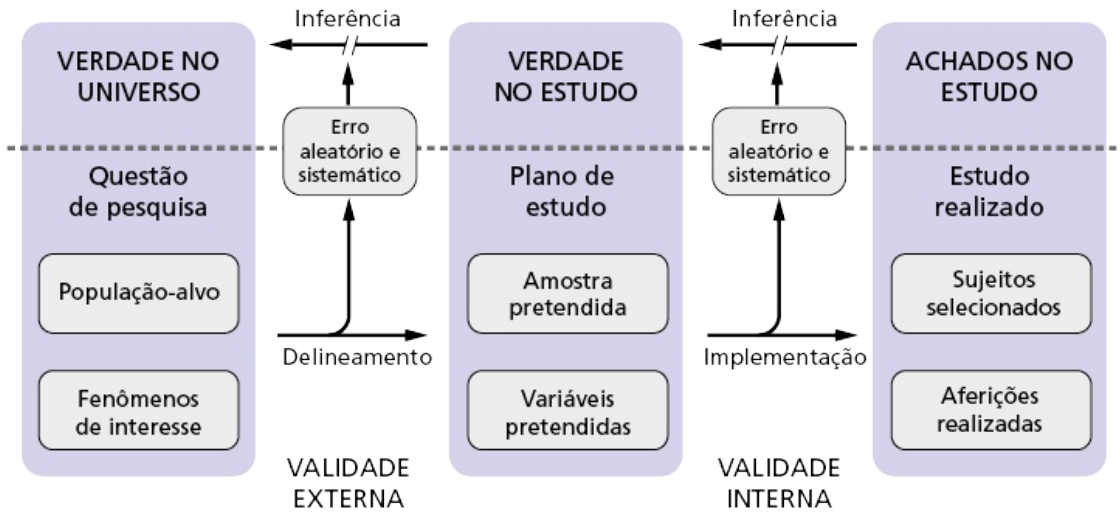



FIGURA 1.5 Fisiologia da pesquisa – como ela funciona.

DELINEANDO O ESTUDO

Desenvolvendo o plano de estudo

O processo de desenvolver um **plano de estudo** inicia com uma **questão de pesquisa** de uma frase que especifica as principais variáveis preditora e de desfecho e a população. A seguir, são produzidas, em sequência, três versões do plano de estudo, cada uma delas maior e mais detalhada do que a anterior.

- **Anteprojeto** ou **pré-proposta** (Tabela 1.1 e Apêndice 1). Esse esboço inicial dos elementos do estudo (1 página) serve como *checklist* padronizado para lembrar o investigador de abordar todos os elementos essenciais. Além disso, a ordem lógica apresentada ajuda a clarear as ideias do investigador sobre o assunto.

- **Protocolo do estudo.** Essa versão ampliada do esboço inicial

geralmente varia de 5 a 15 páginas e é usada para planejar o estudo, encaminhar para aprovação do CEP e solicitar auxílio financeiro. Discutiremos suas partes separadamente ao longo do livro, para então resumi-las no Capítulo 19.

- **Manual de operações.** Esse conjunto de instruções procedimentais, questionários e outros materiais é desenvolvido para garantir uniformidade e padronização com um controle de qualidade adequado (Capítulos 4 e 17).

A questão de pesquisa e o anteprojeto devem ser redigidos nos estágios iniciais da pesquisa. Colocar as ideias no papel ajuda o investigador a transformar as ideias vagas iniciais em planos específicos e fornece bases concretas para buscar sugestões de colaboradores e consultores. Fazê-lo constitui um desafio (é mais fácil falar sobre ideias do que escrever sobre elas), que é, no entanto, recompensado por um início mais rápido e um projeto mais bem elaborado.

O Apêndice 1 é um exemplo de um anteprojeto. Esse resumo de uma página trata mais da anatomia da pesquisa (Tabela 1.1) do que de sua fisiologia (Figura 1.5). Por isso, o investigador deve atentar para os potenciais erros na inferência sobre as aferições na amostra do estudo para os fenômenos de interesse na população. As virtudes e os defeitos de um estudo podem ser revelados explicitando-se as diferenças entre a questão proposta e a questão que o estudo poderá responder. Essa última é consequência dos planos de seleção dos sujeitos e de realização das aferições, bem como dos problemas advindos da implementação do estudo.

Tendo o anteprojeto em mãos e as inferências pretendidas em mente, o investigador pode iniciar o detalhamento de seu protocolo. Isso inclui solicitar sugestões de colaboradores, delinear métodos específicos de recrutamento e aferição, considerar a adequação científica e ética, modificar a questão de pesquisa e o anteprojeto se necessário, pré-testar os métodos específicos de recrutamento e aferição, fazer mudanças adicionais, pedir mais sugestões e assim por diante. Esse processo iterativo é a natureza do delineamento de pesquisa e, como tal, o enfoque adotado no restante deste livro.

Avaliando vantagens e desvantagens

Infelizmente, os erros são parte inerente a todos os estudos. O importante é saber se eles são de magnitude suficiente para mudar substancialmente as conclusões da pesquisa. Ao delinear um estudo, o investigador está na mesma posição que um representante de um sindicato ao negociar um novo contrato. O representante do sindicato começa com uma lista de demandas – redução da jornada de trabalho, aumento de salário, seguro de saúde e assim por diante. Faz, então, concessões, garantindo os itens mais importantes e renunciando àqueles não essenciais ou não realistas. No final das negociações, é fundamental examinar o melhor contrato obtido para decidir se ele não se tornou tão ruim a ponto de não valer a pena assiná-lo.

Os mesmos tipos de concessões devem ser feitas por um investigador ao transformar a questão de pesquisa em um plano de estudo e ao considerar os potenciais problemas de sua implementação. Por um lado, tem-se a questão da validade interna e da validade externa; por outro, tem-se a factibilidade. O último passo vital do negociador do sindicato é frequentemente omitido: o investigador deve decidir se o plano de estudo, como formulado, aborda a questão de pesquisa de forma adequada e se pode ser implementado com níveis aceitáveis de erro. Muitas vezes a resposta é não, e o investigador deve reiniciar o processo. Mas não deve sentir-se desencorajado! Os bons cientistas distinguem-se dos demais não tanto por suas boas ideias de pesquisa, mas pela sua prontidão para rejeitar logo no início aquelas que não irão funcionar e se concentrar naquelas com maior potencial.

RESUMO

1. A **anatomia** da pesquisa é o conjunto de elementos tangíveis que compõem o plano de estudo: a **questão de pesquisa** e sua **relevância**, o **delineamento**, os **sujeitos do estudo** e as **técnicas de aferição**. O desafio é delinear um plano de estudo com elementos que sejam de **custo relativamente baixo** e **fáceis** de implementar.
2. A **fisiologia** da pesquisa é como o estudo funciona. Os achados do estudo são usados para realizar **inferências** sobre o que realmente ocorreu na amostra (**validade interna**) e sobre os eventos do universo

exterior (**validade externa**). O desafio é **delinear** e **implementar** um plano de estudo com um controle adequado sobre as duas maiores ameaças a essas inferências: **erro aleatório** (acaso) e **erro sistemático** (viés).

3. Ao delinear um estudo, pode ser útil considerar a Figura 1.5, que ilustra as relações entre a **questão de pesquisa** (o que o investigador realmente deseja responder no universo), o **plano de estudo** (aquilo que o estudo foi delineado para responder) e o **estudo efetivamente realizado** (aquilo que o estudo irá, de fato, responder, tendo em vista os erros de implementação que podem ser antecipados).
4. Uma boa forma de desenvolver o **plano de estudo** é escrever um resumo de uma frase da **questão de pesquisa** que especifica as variáveis principais e a população e então expandi-lo em um esboço de uma página (**anteprojeto)** que coloca os elementos do estudo em uma sequência padronizada. Posteriormente, o plano de estudo será expandido em um **protocolo** e em um **manual de operações**.
5. O bom **julgamento** por parte do investigador e as sugestões de colaboradores são necessários para **pesar as várias vantagens e desvantagens** envolvidas e determinar a viabilidade geral do projeto.

APÊNDICE 1
Anteprojeto

Este é um anteprojeto que deu origem a um estudo desenvolvido por Valerie Flaherman, MD, MPH, iniciado quando ela ainda era residente de pediatria na Universidade da Califórnia, em São Francisco (UCSF). Para a maioria dos pesquisadores iniciantes, é mais fácil começar com estudos observacionais, mas, neste caso, um ensaio clínico randomizado de tamanho e escopo modestos foi possível, além de ser o único delineamento que poderia responder adequadamente à questão de pesquisa. O artigo de Flaherman et al (1) apresenta os achados, que, se confirmados, podem alterar as orientações vigentes sobre qual é a melhor forma de iniciar o aleitamento materno.

TÍTULO: EFEITO DO INÍCIO PRECOCE DE PEQUENA QUANTIDADE DE FÓRMULA LÁCTEA SOBRE O ALEITAMENTO MATERNO

Questão de pesquisa:

Entre os lactentes que perderam ≥ 5% do peso ao nascer antes de 36 horas de vida, oferecer 10 mL de fórmula láctea por meio de uma seringa após cada mamada antes da produção do leite maduro aumenta a probabilidade de sucesso no aleitamento?

Relevância:

1. A produção de leite materno é baixa antes da produção do leite maduro iniciar, dois a cinco dias após o nascimento.
2. Algumas mães ficam preocupadas se o início da produção do leite maduro é tardia e se o bebê perde muito peso, o que as leva a abandonar o aleitamento materno ainda na primeira semana. Uma estratégia que aumentasse a proporção de mães bem-sucedidas no aleitamento materno poderia ter inúmeros benefícios para a saúde e na esfera psicossocial, tanto para a mãe quanto para o bebê.
3. Estudos observacionais mostraram que a introdução de fórmula nos

primeiros dias após o nascimento está associada com uma diminuição na duração do aleitamento materno. Embora isso poderia ser explicado pelo confundimento por indicação (ver Capítulo 9), esse achado levou as diretrizes da OMS e do CDC a desaconselharem a introdução de fórmula durante a internação pós-parto.

4. Entretanto, uma pequena quantidade de fórmula, associada ao aleitamento materno e ao aconselhamento, poderia tornar a experiência com o aleitamento mais positiva e aumentar a probabilidade de sucesso. Um ensaio clínico randomizado é necessário para avaliar os possíveis benefícios e danos dessa estratégia.

Delineamento:

Ensaio clínico randomizado não cego, porém com avaliação cega dos desfechos

Sujeitos:

- **Critérios de inclusão:** Lactentes saudáveis a termo, com 24 a 48 horas de vida, que perderam ≥ 5% do peso ao nascer nas primeiras 36 horas de vida.
- **Desenho amostral:** Amostra consecutiva de pacientes cujas mães assinaram termo de consentimento informado em dois hospitais universitários no norte da Califórnia.

Variável preditora, alocada aleatoriamente, mas não cega:

- **Controle:** Pais são orientados sobre técnicas para acalmar o bebê.
- **Intervenção:** Pais são orientados a administrar 10 mL de fórmula em uma seringa após cada mamada até o início da produção do leite maduro.

Variáveis de desfecho, avaliadas de forma cega:

1. Qualquer uso de fórmula após uma semana e após um, dois e três meses
2. Qualquer aleitamento materno após uma semana e após um, dois e três meses

3. Nadir do peso

Hipótese nula principal:

Uma quantidade pequena de fórmula nos primeiros dias de vida não afeta a proporção de mulheres que estão amamentando seus bebês após três meses.

REFERÊNCIA

1. Flaherman VJ, Aby J, Burgos AE, et al. Effect of early limited formula on duration and exclusivity of breastfeeding in at-risk infants: an RCT. *Pediatrics*, in press.

1As variáveis preditoras são também conhecidas como **variáveis independentes**, e as de desfecho, como **variáveis dependentes**, porém o significado desses termos é menos evidente, e preferimos evitar seu uso.

CAPÍTULO 2

Elaborando a questão de pesquisa e desenvolvendo o plano de estudo

Steven R. Cummings, Warren S. Browner e Stephen B. Hulley

A **questão de pesquisa** é a incerteza que o investigador pretende resolver ao realizar o estudo. Não faltam questões de pesquisa. Mesmo quando é possível responder a uma questão, continuamos cercados por outras. Por exemplo, ensaios clínicos estabeleceram que tratamentos que bloqueiam a síntese de estradiol (inibidores da aromatase) reduzem o risco de câncer de mama em mulheres que já tiveram câncer em estágio inicial (1). No entanto, isso levou a novas questões: Por quanto tempo deve-se estender o tratamento; o tratamento previne câncer de mama em pacientes com mutações nos genes BRCA1 e BRCA2; e qual a melhor forma de prevenir a osteoporose, que é um efeito adverso conhecido dessas medicações? É possível ir além e elaborar questões primariamente preventivas, como: esses tratamentos são efetivos e seguros para a prevenção do câncer de mama mesmo em mulheres saudáveis?

O desafio de se chegar a uma questão de pesquisa é definir uma questão importante que possa ser transformada em um **plano de estudo** factível e válido. Este capítulo apresenta estratégias para alcançar esse objetivo (Figura 2.1).

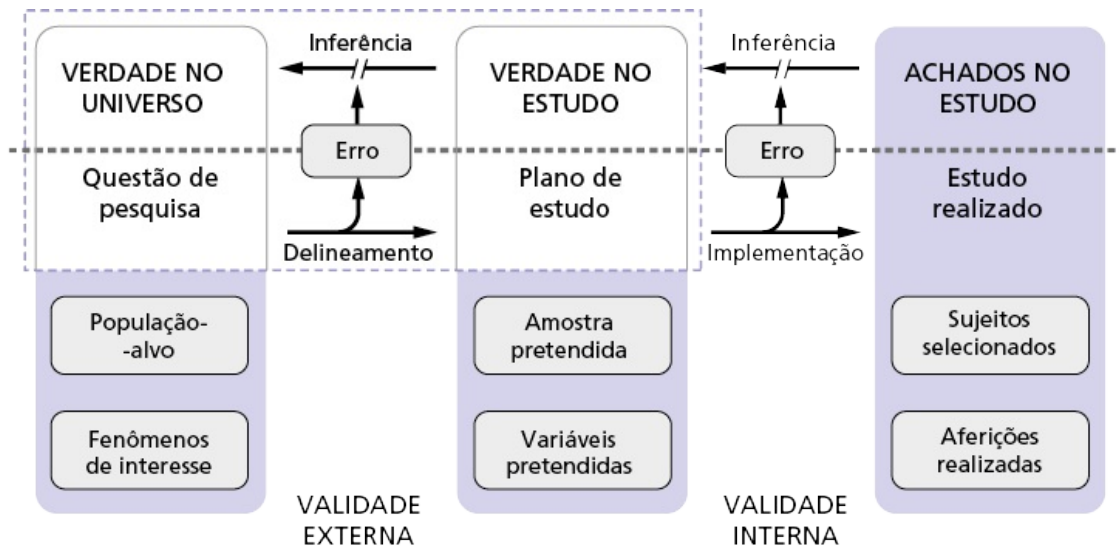



FIGURA 2.1 O foco deste capítulo é a área dentro do retângulo com borda tracejada, o desafio de escolher uma questão de pesquisa que seja de interesse e possa ser transformada em um plano de estudo factível.

ORIGEM DE UMA QUESTÃO DE PESQUISA

Para um investigador experiente, as melhores questões de pesquisa normalmente surgem dos achados e problemas que ele observou em seus **estudos anteriores** e nos de outros pesquisadores da área. O investigador iniciante ainda não teve a oportunidade de desenvolver essa experiência. Embora uma perspectiva nova seja às vezes útil, por permitir que uma pessoa criativa conceba novas abordagens para problemas antigos, a falta de experiência é um impedimento.

Uma boa forma de começar é esclarecer a diferença entre **questão de pesquisa** e **interesse de pesquisa**. Considere a seguinte questão de pesquisa:

- *Participar de sessões de aconselhamento em grupo reduz a probabilidade de violência doméstica em mulheres que recentemente imigraram da América Central?*

Essa questão poderia ter sido formulada por alguém cujo interesse envolve a eficácia do aconselhamento em grupo, ou a prevenção da violência doméstica, ou ainda a melhoria da saúde em imigrantes recentes. Distinguir entre questão de pesquisa e interesse de pesquisa é

importante, porque se pode chegar à conclusão de que a questão de pesquisa não pode ser transformada em um plano de estudo factível, mas o investigador ainda assim poderá satisfazer seu interesse de pesquisa formulando uma questão diferente.

É claro que é impossível formular uma questão de pesquisa se você não tiver clareza sobre seu interesse de pesquisa (além de saber que você precisa ter um). Se você se encontrar nesse barco, saiba que não está sozinho. Muitos investigadores iniciantes ainda não descobriram assuntos que os interessem e que possam ser transformados em um plano de estudo factível de ser implementado por eles. Você pode começar considerando que tipos de artigos chamaram sua atenção em uma revista médica. Ou é possível que você tenha se sentido incomodado com relação a um paciente específico cujo tratamento você considerou inadequado. O que poderia ter sido feito de forma diferente e que poderia ter melhorado seu desfecho? Ou é possível que um dos seus preceptores tenha dito que a hipocaliemia causa muita sede, ao passo que outro disse exatamente o contrário, de forma igualmente dogmática.

Dominando a literatura

É importante o **domínio da literatura** publicada em uma determinada área de estudo, o que constitui um precursor necessário para uma boa pesquisa. O investigador iniciante deve fazer uma busca ampla de toda a literatura publicada nas áreas pertinentes à questão de pesquisa e ler criticamente artigos originais importantes. Uma **revisão sistemática** é um bom próximo passo para desenvolver domínio sobre o assunto. A revisão da literatura pode servir como fonte de informações na solicitação de auxílio financeiro e na preparação de relatórios de pesquisa. É importante lembrar que avanços recentes podem ser do conhecimento de investigadores ativos em uma determinada área muito antes de serem publicados. Portanto, para dominar um assunto, é necessário participar de congressos e estabelecer vínculos com **especialistas** na área.

Estando alerta a novas ideias e técnicas

Além de revisar a literatura em busca de questões de pesquisa, é útil comparecer a **congressos** onde são apresentados trabalhos recentes da área. A discussão sobre os trabalhos apresentados pode ser suplementada

por conversas informais com outros cientistas nas sessões de pôsteres e durante o intervalo. Um investigador iniciante que vence a timidez e consegue conversar com um palestrante durante o *coffee break* poderá ver que essa experiência é muito proveitosa e, por vezes, acabará atraindo um novo colaborador sênior nesse processo. Quando se sabe com antecedência da presença de um palestrante especialmente relevante, vale a pena contatá-lo previamente para agendar um encontro durante o congresso.

Uma **atitude cética** sobre ideias correntes pode estimular boas questões de pesquisa. Por exemplo, acreditava-se que ferimentos cutâneos que atravessavam a derme exigiam aproximação com suturas para assegurar uma cicatrização rápida e um resultado estético satisfatório. No entanto, Quinn e colaboradores observaram, mediante experiência pessoal e séries de casos, que os ferimentos cutâneos de tamanho moderado sofrem reparo adequado independentemente de seus bordos terem sido aproximados ou não (2). Eles realizaram um ensaio clínico randomizado no qual todos os pacientes com lacerações na mão com menos de 2 cm de comprimento receberam irrigação com água de torneira e um curativo com antibióticos por 48 horas. Um dos grupos foi alocado aleatoriamente para receber sutura e o outro grupo para não receber sutura. O grupo em que a sutura foi feita teve um tratamento mais doloroso e que consumiu mais tempo no serviço de emergência, porém uma avaliação cega mostrou que o tempo até a cicatrização e os resultados estéticos foram semelhantes aos do grupo em que a sutura não foi feita. Essa conduta agora se tornou padrão na prática clínica.

A adoção de **novas tecnologias** gera novas percepções e questionamentos sobre problemas clínicos bem-conhecidos, o que, por sua vez, pode originar novos paradigmas (3). Por exemplo, os avanços de imagem e de tecnologias moleculares e genéticas originaram pesquisas clínicas translacionais que levaram a novos tratamentos e exames que, por sua vez, mudaram a prática clínica. Da mesma forma, aplicar um novo conceito, tecnologia ou achado de uma área a um problema de outra área pode gerar boas questões de pesquisa. Por exemplo, a baixa densidade óssea é um fator de risco para fraturas. Investigadores aplicaram essa tecnologia para outros desfechos, demonstrando que mulheres com baixa densidade óssea têm maiores taxas de declínio cognitivo (4), estimulando

pesquisas sobre fatores, como baixos níveis endógenos de estrogênio, que poderiam levar à perda tanto de massa óssea quanto de memória.

Mantendo a imaginação ativa

A **observação** cuidadosa de pacientes tem gerado muitos estudos descritivos e é fonte frutífera de questões de pesquisa. A **docência** também é uma fonte excelente de inspiração; ideias de estudos costumam surgir na preparação de aulas ou durante discussões com alunos questionadores. Como em geral não há tempo suficiente para desenvolver essas ideias imediatamente, é útil registrá-las em um **arquivo de computador** ou em um caderno para futura referência.

A **criatividade** exerce papel fundamental na concepção de questões de pesquisa, na proposição de novos métodos para responder a questões antigas e no prazer em jogar com as ideias. Algumas ideias criativas surgem durante conversas informais com colaboradores no horário de almoço; outras surgem em reuniões de pequenos grupos nas quais se discutem pesquisas recentes ou as ideias dos próprios integrantes do grupo. Muitas inspirações vêm do indivíduo, enquanto ele prepara uma palestra, toma um banho, navega na internet ou simplesmente senta e pensa sobre o assunto. O medo de ser criticado ou de soar estranho pode abortar precocemente ideias novas. O truque é não forçar, visualizar claramente um problema não resolvido e então ligar a tomada mental que deixa a mente correr livremente em torno dele. É preciso também **persistência** para retornar a um problema complicado tantas vezes quanto necessário até alcançar uma resolução.

Escolhendo e trabalhando com um mentor

Nada substitui a experiência pessoal na orientação dos múltiplos julgamentos envolvidos na concepção de uma questão de pesquisa e no desenvolvimento de um plano de estudo. Portanto, uma estratégia essencial para o jovem investigador é tornar-se aprendiz de um **mentor** experiente que tenha tempo e interesse em trabalhar com ele regularmente.

Um bom mentor estará disponível para encontros regulares e discussões informais e irá encorajar ideias criativas, contribuir com a sabedoria que só vem com a experiência, ajudar a assegurar tempo disponível para

pesquisa, abrir portas para a formação de redes de pesquisa e oportunidades de financiamento, encorajar o desenvolvimento de trabalhos independentes e colocar o nome do novo investigador em primeiro lugar em auxílios e publicações sempre que for conveniente. Às vezes é desejável ter mais de um mentor, representando diferentes áreas do conhecimento. Um bom relacionamento desse tipo também pode levar a recursos dos quais um investigador jovem necessita – espaço de escritório, acesso a pacientes, bancos de dados e bancos de amostras, laboratórios especializados, recursos financeiros e uma equipe de pesquisa.

Um mentor ruim, por outro lado, pode ser uma barreira. Um mentor pode comprometer a carreira do jovem investigador, por exemplo, ao assumir o crédito pelos achados que surgem do trabalho de seu orientando ou ao assumir o papel principal ao publicar ou apresentar os dados. Com frequência, muitos mentores estão simplesmente ocupados ou distraídos demais para prestar atenção às necessidades do investigador iniciante. Seja qual for a situação, a partir do momento em que as discussões com o mentor se mostrarem infrutíferas, recomendamos buscar uma forma de partir para um supervisor mais apropriado, possivelmente pedindo o auxílio de um colaborador sênior para auxiliar na negociação. A **troca de mentor** pode ser traumática, motivo pelo qual enfatizamos a importância de escolher um bom mentor desde o início; esta é possivelmente *a decisão mais importante* que um investigador iniciante deve tomar.

O seu mentor pode lhe dar acesso a um banco de dados e pedir que você desenvolva uma questão de pesquisa. Nesse caso, é importante identificar (1) se os dados disponíveis atendem aos seus próprios interesses de pesquisa e (2) a qualidade do banco de dados. Se os dados não atenderem aos seus interesses de pesquisa ou forem de má qualidade, é melhor buscar um outro projeto.

CARACTERÍSTICAS DE UMA BOA QUESTÃO DE PESQUISA

Uma questão de pesquisa, para levar a um bom plano de estudo, deve ser: factível, interessante, nova (original, inovadora), ética e relevante (formando o acrônimo **FINER**; Tabela 2.1).

TABELA 2.1 Critérios FINER para uma boa questão de pesquisa e um bom plano de estudo

Factível

Número adequado de sujeitos
Domínio técnico adequado
Viável em termos de tempo e custos
Escopo manejável
Capaz de obter financiamento

Interessante

Descobrir a resposta é estimulante para o investigador e para seus colaboradores?
Nova (original, inovadora)
Fornece novos achados
Confirma, refuta ou expande achados anteriores
Pode levar a inovações em conceitos sobre saúde e doença, na prática médica ou em metodologias de pesquisa

Ética

Adequada para um estudo que um comitê de ética em pesquisa aprovaria

Relevante

Provavelmente terá impacto significativo no conhecimento científico, na prática clínica ou nas políticas de saúde
Pode influenciar direcionamentos futuros de pesquisa

Factível

É fundamental conhecer desde o início os limites e problemas práticos de se estudar uma questão de pesquisa antes de se despender muito tempo e esforço em um caminho que se mostrará inviável.

- **Número de sujeitos.** Muitos estudos não alcançam os objetivos pretendidos por não conseguirem um número suficiente de sujeitos. O primeiro passo é fazer uma estimativa preliminar das exigências de tamanho de amostra do estudo (Capítulo 6). O passo seguinte é estimar o número de sujeitos com possibilidade de estarem disponíveis para o estudo, o número que seria excluído ou não aceitaria participar e o número de sujeitos que seriam perdidos no seguimento. Mesmo um planejamento cuidadoso costuma produzir estimativas excessivamente otimistas, e o investigador deve ter certeza de que há um número suficiente de sujeitos elegíveis e dispostos a participar. Às vezes é necessário realizar um levantamento-piloto ou revisão de prontuários para se assegurar disso. Se o número de sujeitos aparenta ser insuficiente, pode-se considerar uma série de estratégias. Elas incluem expandir os critérios de inclusão, eliminar os critérios de exclusão desnecessários, aumentar o prazo para arrolamento de sujeitos, obter fontes adicionais de sujeitos, desenvolver abordagens de medição mais

precisas, convidar novos colaboradores para trabalhar no projeto como um estudo multicêntrico e mudar o delineamento do estudo.

- **Domínio técnico.** Os investigadores devem ter as habilidades, o equipamento e a experiência necessários para delinear o estudo, recrutar os sujeitos, medir as variáveis e gerenciar e analisar os dados. Os consultores podem ajudar nos aspectos técnicos que os investigadores não dominam, mas para as áreas importantes do estudo é importante ter um colaborador experiente com maior envolvimento no estudo no papel de um coinvestigador. Por exemplo, é geralmente bom incluir um estatístico como membro do grupo de pesquisa desde o início do processo de planejamento. A estratégia mais fácil é usar abordagens familiares e estabelecidas, pois o processo de desenvolver novos métodos e habilidades demanda tempo e é de resultado incerto. Quando for necessário desenvolver uma abordagem nova específica, como, por exemplo, a medição de um novo biomarcador, é preciso obter os conhecimentos técnicos sobre como incorporar essa inovação.

- **Tempo e custos envolvidos.** É importante estimar os custos de cada componente do projeto, tendo em mente que o tempo e os recursos financeiros necessários geralmente irão exceder o planejado. Se os custos são proibitivos, as únicas opções são considerar um delineamento menos dispendioso ou buscar fontes adicionais de financiamento. Se o estudo for muito caro ou demandar muito tempo, é melhor sabê-lo de antemão, quando a questão ainda pode ser modificada ou abandonada, antes de despender muito esforço.

- **Escopo.** Frequentemente surgem problemas quando um investigador tenta fazer demais, realizando muitas medições em ocasiões repetidas em um número grande de sujeitos, em uma tentativa de responder a muitas questões de pesquisa. A solução é reduzir o escopo do estudo e focar somente os objetivos mais importantes. Pode ser difícil desistir da oportunidade de responder a questões secundárias interessantes, mas o benefício pode ser uma resposta melhor para a questão principal.

- **Possibilidade de obter financiamento.** Poucos investigadores têm os

recursos pessoais ou institucionais para financiar seus próprios projetos de pesquisa, especialmente se for necessário arrolar e seguir os participantes, ou se for necessário realizar aferições de alto custo. Mesmo o protocolo de pesquisa mais bem elaborado não será factível se não houver financiamento para o estudo. A busca por fontes de financiamento é discutida no Capítulo 19.

Interessante

O investigador pode ter muitas motivações para se aventurar em uma determinada questão de pesquisa: porque poderá conseguir suporte financeiro, porque é uma nova etapa lógica ou importante na construção de uma carreira, ou porque chegar à verdade sobre o assunto parece interessante. Apraz-nos essa última motivação; ela cresce à medida que é exercida e fornece a intensidade de esforço necessária para superar os vários obstáculos e frustrações envolvidos no processo. No entanto, é melhor assegurar-se de que o investigador não seja a única pessoa entusiasmada com uma determinada questão de pesquisa. Deve-se confirmar o interesse sobre uma questão com mentores e especialistas de fora do grupo e com representantes de potenciais fontes de financiamento, antes de despender muita energia para desenvolver um plano de pesquisa ou proposta de auxílio que os pares e as agências de financiamento poderão considerar de pouca importância.

Nova (original, inovadora)

Boas pesquisas clínicas produzem informações novas. Um estudo que meramente reitera o que já foi estabelecido não vale o esforço e os recursos despendidos e é improvável que obtenha financiamento. A potencial novidade do estudo proposto pode ser avaliada revisando amplamente a literatura, consultando especialistas que conhecem as pesquisas em andamento e ainda não publicadas e buscando resumos de projetos na sua área de interesse que já foram financiados, por meio do portal do National Institutes of Health (NIH) Research Portfolio Online Reporting Tools (**RePORT**, http://report.nih.gov/categorical_spending.aspx.).[1] As revisões dos estudos submetidos ao NIH dão grande peso para se o estudo proposto é **inovador** (5), ou seja, se os resultados poderão mudar paradigmas da

pesquisa ou da prática clínica por meio do uso de novos conceitos, métodos ou intervenções (Capítulo 19). Embora o caráter inovador seja um critério importante, a questão não precisa ser totalmente original. Pode valer a pena questionar se uma observação anterior pode ser replicada, se os achados em uma população se aplicam a outra ou se um novo método de aferição pode esclarecer a relação entre fatores de risco conhecidos e uma doença. Um estudo confirmatório é especialmente útil se evitar as limitações de estudos anteriores ou se o resultado a ser confirmado for inesperado.

Ética

Uma boa questão de pesquisa deve ser ética. Se o estudo impõe riscos físicos ou invasão de privacidade inaceitáveis (Capítulo 14), o investigador deve buscar outras formas de responder à questão. Caso haja incerteza sobre se um estudo é ético, é importante discuti-lo em um estágio inicial com um representante do Comitê de Ética em Pesquisa (CEP) da sua instituição.

Relevante

Uma forma de decidir sobre a relevância é imaginar os vários desfechos que poderiam ocorrer e considerar como cada um poderia trazer avanços para o conhecimento científico, influenciar o manejo clínico e as políticas de saúde ou direcionar pesquisas futuras. As agências de fomento costumam colocar grande ênfase na **relevância** de um estudo proposto, ou seja, na importância do problema, em como o projeto irá aprimorar o conhecimento científico e em como o resultado irá mudar conceitos, métodos ou serviços clínicos.

DESENVOLVENDO A QUESTÃO DE PESQUISA E O PLANO DE ESTUDO

Vale a pena colocar no papel, em um estágio preliminar, a questão de pesquisa e o **anteprojeto** (esboço do plano de estudo, de 1 página) (Apêndice 1). Isso exige uma certa autodisciplina, mas obriga o investigador a tornar mais claras suas ideias sobre o plano de estudo e a identificar problemas específicos que podem demandar mais atenção. O

anteprojeto também fornece uma base para os colaboradores poderem contribuir com sugestões específicas.

Problemas e abordagens

Duas abordagens complementares para lidar com problemas que surgem no desenvolvimento de uma questão de pesquisa merecem especial atenção.

A primeira é buscar um bom **aconselhamento**. Para tanto, recomenda-se incluir no grupo de pesquisa representantes de cada um dos principais aspectos do estudo, incluindo ao menos um pesquisador sênior. Além disso, é aconselhável consultar especialistas para a localização de pesquisas sobre o assunto e para a escolha e definição das técnicas de medição. Em certas ocasiões, um **especialista local** será suficiente, mas muitas vezes é útil contatar pesquisadores de outras instituições que publicaram trabalhos pertinentes ao assunto estudado. Um investigador iniciante pode sentir-se intimidado pela ideia de escrever ou ligar para alguém que conhece apenas como autor no *Journal of the American Medical Association*, mas a maioria dos cientistas responde de forma favorável a esses pedidos.

A segunda abordagem é permitir que o plano de estudo se desenvolva gradualmente por meio de um **processo iterativo**, fazendo alterações incrementais no delineamento do estudo, estimando o tamanho de amostra, revisando o projeto com colaboradores, realizando pré-testes dos pontos principais do estudo e revisando o conjunto. Uma vez escrito o anteprojeto, os colaboradores farão uma revisão formal que geralmente irá resultar em melhorias importantes. À medida que o protocolo vai tomando forma, estudos pilotos sobre disponibilidade e disposição de um número suficiente de sujeitos participantes poderão levar a alterações no plano de recrutamento. O exame de imagem preferido pode ter custo proibitivo, exigindo a busca de uma alternativa menos dispendiosa.

Questões principais e secundárias

Muitos estudos têm mais de uma questão de pesquisa. Os ensaios clínicos, em geral, abordam o efeito da intervenção sobre mais de um desfecho. Por exemplo, o Women’s Health Initiative foi delineado para determinar se a redução da ingestão de gordura alimentar poderia reduzir o risco de

câncer de mama, mas uma questão secundária importante era avaliar o efeito sobre eventos coronarianos (5). Quase todos os estudos de coorte e de caso-controle examinam vários fatores de risco para cada desfecho. A vantagem de um delineamento com várias questões de pesquisa é a eficiência resultante, pois várias respostas poderão surgir de um único estudo. As desvantagens são o aumento de complexidade do delineamento e da implementação do estudo e das inferências estatísticas a partir de múltiplas hipóteses (Capítulo 5). Uma estratégia sensata é determinar **uma única questão de pesquisa principal** que norteará o desenvolvimento do plano de estudo e a estimativa do tamanho de amostra. Essa questão pode ser suplementada por **questões de pesquisa secundárias** sobre outros preditores ou desfechos que também poderão trazer conclusões de valor.

PESQUISA TRANSLACIONAL

A **pesquisa translacional** preocupa-se em como traduzir achados da torre de marfim da academia para o "mundo real" e como assegurar-se de que a criatividade científica tenha um impacto favorável na saúde pública. Existem dois tipos de pesquisa translacional (6) (Figura 2.2):

- Aplicação de achados da pesquisa básica em estudos clínicos com pacientes (às vezes abreviado como **T1**), e
- Aplicação de achados desses estudos para alterar práticas de saúde na comunidade (às vezes abreviado como **T2**).

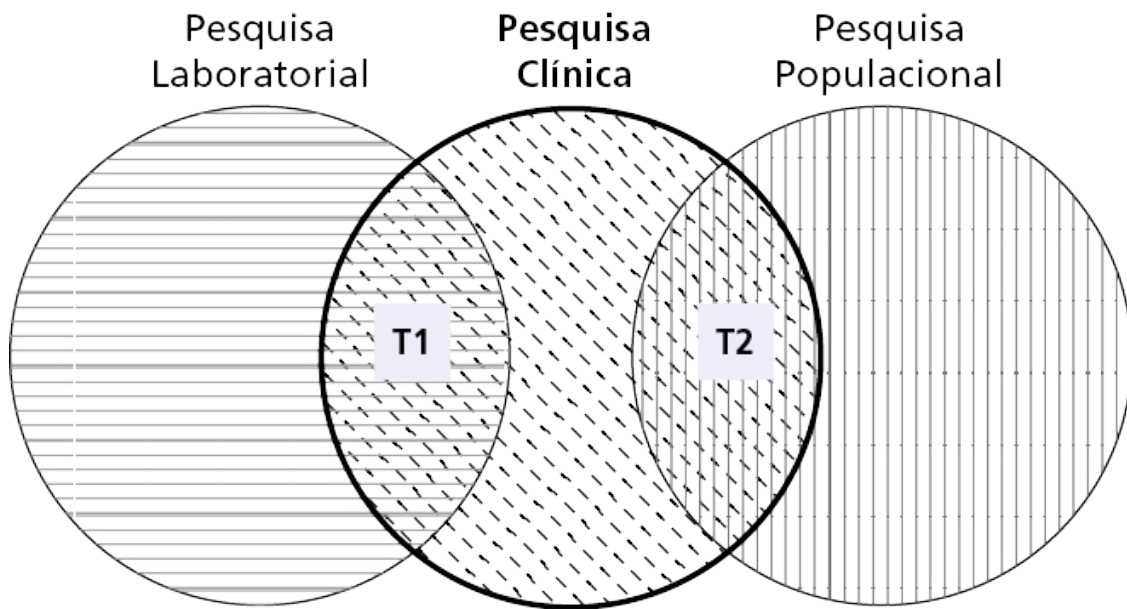



FIGURA 2.2 A pesquisa translacional é o componente da pesquisa clínica que interage com a pesquisa básica (área hachurada T1) ou com a pesquisa populacional (área hachurada T2).

Ambas as formas de pesquisa translacional requerem a identificação de um achado de pesquisa para ser "traduzido". Assim como um tradutor literário precisa primeiro identificar o romance ou poema que irá traduzir, para realizar uma pesquisa translacional deve-se primeiro concentrar as atenções em um achado científico ou em uma tecnologia nova que poderá ter um impacto importante na pesquisa ou prática clínica ou em políticas de saúde pública. Para auxiliar nessa escolha desafiadora, pode ser útil prestar atenção aos seus colegas pesquisadores quando estiverem conversando sobre os últimos achados de suas pesquisas, às apresentações sobre métodos novos em congressos e a especulações na literatura sobre mecanismos.

Traduzindo pesquisas do laboratório para a prática clínica (Pesquisas T1)

Uma série de **ferramentas** se tornaram disponíveis para investigações clínicas, incluindo sequenciamento genético, *arrays* de expressão gênica, imagem molecular e proteômica. Do ponto de vista de um pesquisador clínico, não há nada epidemiologicamente diferente sobre essas aferições, tecnologias ou resultados de testes inovadores. O capítulo sobre aferições será útil para planejar estudos envolvendo esses tipos de medidas

(Capítulo 4); também serão úteis as orientações sobre delineamentos de estudos (Capítulos 7-12), amostragem (Capítulo 3) e tamanho de amostra (Capítulo 6). Um aspecto especialmente relevante para a genômica e outras "ômicas" é a preocupação quanto aos testes de hipóteses múltiplas (Capítulo 5).

No entanto, em comparação com a pesquisa clínica tradicional, para ser um investigador translacional T1 de sucesso, muitas vezes é necessário ter habilidades adicionais ou trabalhar com um colaborador que tenha essas habilidades. A **pesquisa da bancada ao leito** requer um conhecimento profundo sobre a ciência básica envolvida. Embora muitos pesquisadores clínicos acreditem que possam dominar esse conhecimento – assim como muitos pesquisadores da área básica acreditam que a pesquisa clínica não requer nenhum treinamento especial –, na prática, as habilidades envolvidas são muito diferentes. Por exemplo, suponha que um cientista da área básica tenha identificado um gene que afeta o ritmo circadiano em ratos. Um **investigador clínico** que tem experiência na área do sono tem acesso a um estudo de coorte com dados sobre os ciclos de sono e um banco com DNA armazenado e deseja estudar se existe associação entre variantes no homólogo humano desse gene e padrões de sono. Para propor um estudo T1 sobre essa associação, é preciso buscar colaboradores familiarizados com o gene envolvido, bem como com as vantagens e limitações dos vários métodos de genotipagem.

Agora imagine que um **investigador de laboratório** tenha descoberto um padrão singular de expressão gênica em amostras de biópsias de pacientes com câncer de mama. Ele não deveria propor um estudo para avaliar o uso desse padrão como teste para predizer o risco de recorrência de câncer de mama sem colaborar com alguém que conheça bem a importância de aspectos próprios da pesquisa clínica, como confiabilidade teste-reteste, amostragem e cegamento. Esse colaborador deveria também ter uma boa compreensão sobre os efeitos da probabilidade prévia de doença sobre a aplicabilidade da descoberta. Uma boa pesquisa translacional requer conhecimento sobre mais de uma área. Assim, por exemplo, uma equipe de pesquisa interessada em testar um novo medicamento pode precisar ter cientistas familiarizados com biologia molecular, farmacocinética, farmacodinâmica, ensaios clínicos de Fase I e II e com padrões da prática clínica no campo da medicina pertinente ao

medicamento.

Traduzindo a pesquisa clínica para a pesquisa populacional (T2)

Para conduzir estudos que tentam aplicar os achados de ensaios clínicos para **populações maiores e mais diversificadas**, é preciso saber identificar grupos de alto risco ou negligenciados, compreender a diferença entre rastreamento e diagnóstico e saber implementar mudanças nos sistemas de saúde. Do ponto de vista prático, esse tipo de pesquisa geralmente requer acesso a grandes grupos de pacientes (ou médicos), como aqueles conveniados a planos de saúde ou a clínicas de grande porte. O apoio e aconselhamento do chefe do departamento, de um chefe de serviço de um hospital afiliado, do coordenador de um serviço de medicina de grupo ou de um representante de uma organização comunitária podem ser úteis ao planejar esses estudos.

Alguns investigadores tentam um atalho ao conduzir estudos translacionais desse tipo, expandindo um estudo em sua própria clínica para incluir também pacientes de seus colegas de trabalho (p. ex., em um ambulatório de um hospital acadêmico), em vez de buscar médicos da comunidade. Essa prática é semelhante a traduzir uma peça do Aristófanes para o grego moderno – continuará não sendo muito útil para leitores de língua inglesa. O Capítulo 18 enfatiza a importância de se estender o alcance da pesquisa o máximo possível para dentro da comunidade.

Ao avaliar a aplicação de resultados de uma pesquisa para populações maiores, muitas vezes é necessário adaptar os métodos de modo que eles se adequem a organizações. Por exemplo, em um estudo para saber se um novo programa de dieta e de exercício físico focado na unidade de saúde terá efeito na comunidade, pode não ser possível alocar separadamente pacientes individuais para as intervenções. Uma solução seria alocar aleatoriamente serviços de saúde. Para isso, pode ser necessário consultar algum especialista em amostragem por conglomerados e nas análises estatísticas envolvidas. Muitos projetos de pesquisa T2 que buscam melhorar os cuidados médicos usam como desfechos variáveis substitutas de “processo”. Por exemplo, se ensaios clínicos já demonstraram que um novo tratamento reduz a mortalidade por sepse, um estudo translacional comparando dois programas para implementar e promover o uso do novo

tratamento não precisaria necessariamente usar a mortalidade como desfecho. Ele poderia simplesmente comparar as porcentagens de pacientes com sepse que receberam o novo tratamento. Partir da pesquisa em ambientes que foram planejados com o objetivo de realizar pesquisa para organizações que foram planejadas para cuidados médicos ou outros objetivos requer flexibilidade e criatividade em aplicar princípios que assegurem o máximo de rigor e validade possíveis para o estudo.

RESUMO

1. Todos os estudos partem de uma **questão de pesquisa** que aborda aquilo que o investigador gostaria de saber. A meta é encontrar uma questão relevante que possa ser desenvolvida em um bom **plano de estudo**.
2. O **domínio do assunto** é essencial para desenvolver questões de pesquisa que mereçam ser estudadas. Uma **revisão sistemática** de estudos pertinentes a uma área de interesse é um bom lugar para iniciar. Comparecer a **conferências** e ficar alerta a novos resultados expande os conhecimentos do investigador para além do que já foi publicado.
3. A decisão mais importante para o iniciante é escolher um ou dois pesquisadores seniores para atuarem como seus **mentores**, ou seja, investigadores experientes que dedicarão parte do seu tempo para se **reunir**, oferecer **recursos** e **contatos**, encorajar a **criatividade** e promover a **independência** e visibilidade dos jovens cientistas que eles estão supervisionando.
4. Boas questões de pesquisa surgem da identificação de novos colaboradores em **conferências**, de um pensamento crítico sobre a prática clínica e seus problemas, da aplicação de **novos métodos** a questões antigas e da reflexão sobre ideias que surgem das atividades de **ensino**, dos **devaneios** e da **busca persistente** por soluções para problemas ainda não resolvidos.
5. Antes de dedicar muito tempo e esforço para escrever uma proposta ou realizar um estudo, o investigador deve avaliar se a questão de pesquisa e o plano de estudo são FINER: **factíveis, interessantes, novos** (inovadores, originais)**, éticos** e **relevantes**. As agências de

financiamento de pesquisa costumam priorizar propostas inovadoras que tenham **impactos significativos** na ciência e na saúde.

6. Em um estágio inicial, a questão de pesquisa deve ser escrita em um **esboço do estudo** de uma página que descreva quantos sujeitos serão necessários, como eles serão selecionados e que aferições serão feitas.
7. Desenvolver a questão de pesquisa e o plano de estudo é um **processo iterativo** que envolve conversas com consultores e amigos, familiaridade com a literatura e estudos-piloto de abordagens de recrutamento e medição.
8. A maior parte dos estudos tem mais de uma questão, e é aconselhável focar em **uma única questão principal** ao delinear e implementar o estudo.
9. A **pesquisa translacional** é um tipo de pesquisa clínica que estuda como aplicar os achados da ciência básica para estudos com pacientes (**T1**) e, então, como aplicar esses achados para melhorar as práticas de saúde na comunidade (**T2**); requer a colaboração de investigadores com experiência em pesquisa de **laboratório** e investigadores com experiência em **estudos populacionais**, usando os **métodos de pesquisa clínica** apresentados neste livro.

REFERÊNCIAS

1. The ATAC Trialists Group. Anastrazole alone or in combination with tamoxifen *versus* tamoxifen alone for adju-vant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomized trials. *Lancet* 2002;359:2131–2139.
2. Quinn J, Cummings S, Callaham M, et al. Suturing *versus* conservative management of lacerations of the hand: randomized controlled trial. *BMJ* 2002;325:299–301.
3. Kuhn TS. *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press, 1962.
4. Yaffe K, Browner W, Cauley J, et al. Association between bone mineral density and cognitive decline in older women. *J Am Geriatr Soc* 1999;47:1176–1182.
5. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer. *JAMA* 2006;295:629–642.
6. Zerhouni EA. US biomedical research: basic, translational and clinical sciences. *JAMA* 2005;294: 1352–1358.

1 N. de R.T. No Brasil, pode ser verificado por meio da Plataforma Brasil (http://aplicacao.saude.gov.br/plataformabrasil/login.jsf).

CAPÍTULO 3

Escolhendo os sujeitos do estudo: especificação, amostragem e recrutamento

Stephen B. Hulley, Thomas B. Newman e Steven R. Cummings

Escolher corretamente os sujeitos do estudo é fundamental para garantir que os resultados representem fielmente o que ocorre na **população** de interesse. Para tanto, o protocolo deve especificar uma **amostra** de sujeitos que possam ser estudados dentro das limitações de orçamento e de tempo disponível (isto é, que não seja nem muito grande nem de difícil acesso) e que seja, ao mesmo tempo, suficientemente grande para controlar o erro aleatório e representativa para permitir a generalização dos achados para a população que se quer estudar. A **capacidade de generalização** raramente é uma simples questão de sim ou de não; envolve, na verdade, um julgamento qualitativo complexo que depende da escolha da população e do desenho amostral.

Trataremos da escolha do *número* apropriado de sujeitos para estudo no Capítulo 6. Neste capítulo, abordaremos o processo de **especificação** e **amostragem** de sujeitos representativos da população e viáveis de serem estudados (Figura 3.1). Discutiremos também estratégias para o **recrutamento** dessas pessoas.

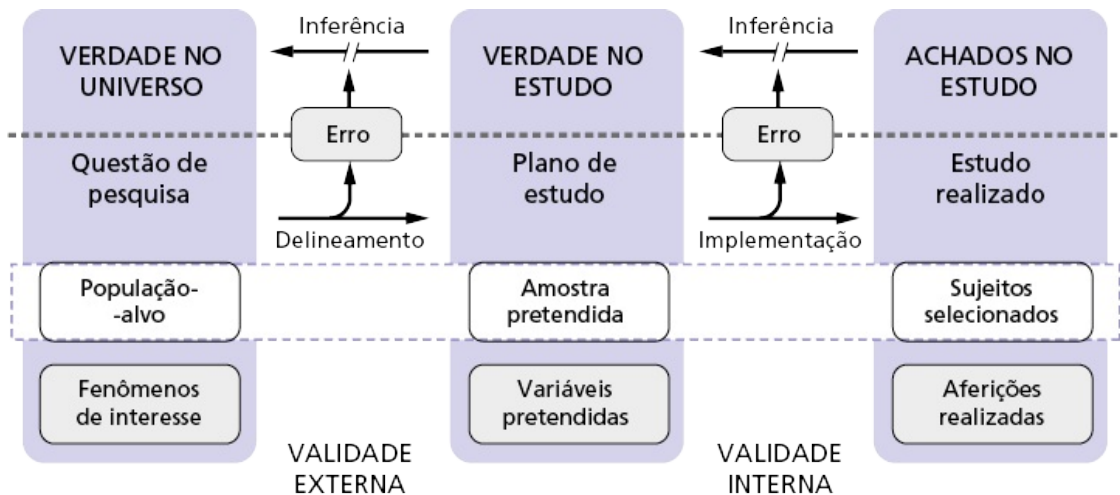



FIGURA 3.1 Este capítulo tem como foco a escolha de uma amostra de sujeitos para o estudo que representem a população de interesse para a questão de pesquisa.

TERMOS E CONCEITOS BÁSICOS

Populações e amostras

População é um conjunto completo de pessoas que apresentam determinadas características em comum, e **amostra** é um subconjunto da população. No uso leigo dessas expressões, as características que definem uma população tendem a ser geográficas; fala-se, por exemplo, da população do Canadá. Em pesquisa, essas características são também clínicas, demográficas e temporais:

- Características clínicas e demográficas definem a **população-alvo**, o conjunto maior de pessoas ao redor do mundo para as quais os resultados poderão ser generalizados (p. ex., *adolescentes com asma*).
- A **população acessível** é um subconjunto geográfica e temporalmente bem definido da população-alvo disponível para estudo (p. ex., *adolescentes com asma que atualmente moram na cidade do investigador*).
- A **amostra pretendida do estudo** é o subconjunto da população acessível que o investigador quer incluir no estudo.

- A **amostra real do estudo** é o grupo de sujeitos que realmente participam do estudo.

Generalizando os resultados do estudo

O estudo clássico de Framingham foi uma abordagem pioneira para delinear de forma científica um estudo que permitiria fazer **inferências** a partir de achados observados em uma amostra para uma população (Figura 3.2).

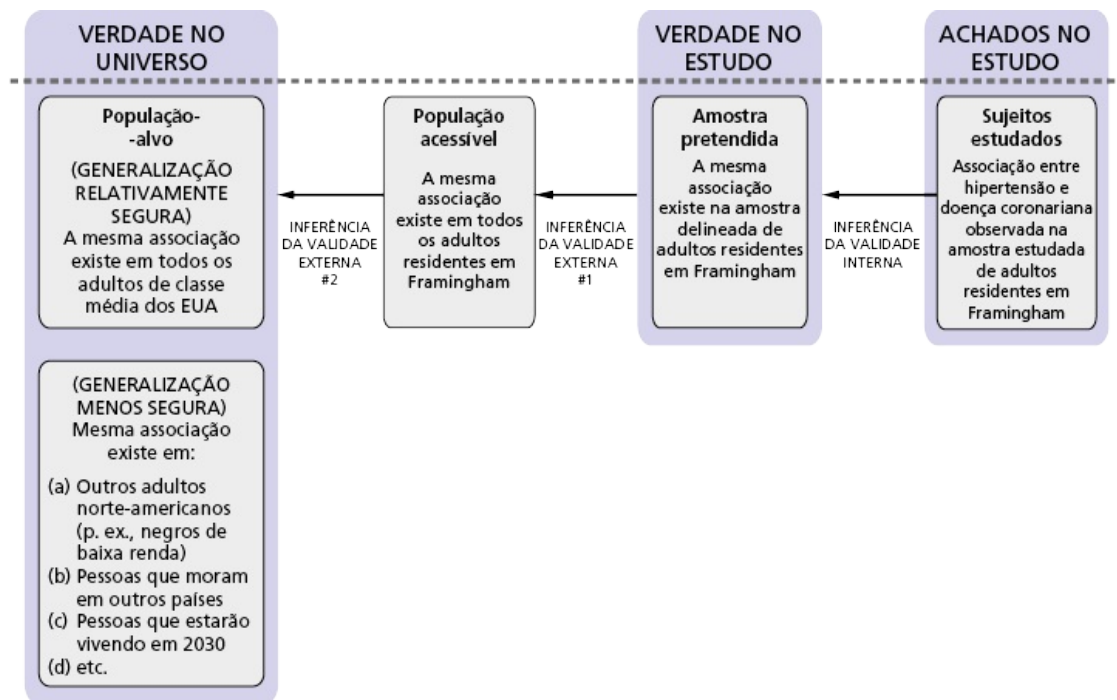



FIGURA 3.2 As inferências na generalização a partir dos sujeitos do estudo para as populações-alvo ocorrem da direita para a esquerda.

A estratégia de amostragem previa a identificação de todas as famílias em Framingham com pelo menos uma pessoa na faixa etária dos 30 aos 59 anos, sendo todas as famílias listadas por ordem de endereço. Então, as pessoas dessa faixa etária nas primeiras duas famílias de cada conjunto de três famílias eram convidadas a participar. Essa estratégia de amostragem "sistemática" é mais suscetível a erros do que selecionar cada sujeito por meio de um processo aleatório (como discutido mais adiante neste capítulo), porém dois problemas mais graves foram o fato de que um terço dos moradores de Framingham selecionados para o estudo se recusou a

participar e que, no lugar deles, os investigadores aceitaram moradores nessa faixa etária que não estavam na amostra selecionada e se voluntariaram para participar no estudo (1).

Visto que as pessoas as quais aceitam participar são em geral mais saudáveis do que as que não aceitam, principalmente se forem voluntárias, as características da amostra real certamente diferem das da amostra pretendida. Nenhuma amostragem é livre de erros; o essencial, no entanto, é saber se eles são de magnitude suficiente para levar a uma resposta incorreta para a questão de pesquisa. Os erros de amostragem desse estudo não parecem ser grandes o suficiente para invalidar a conclusão de que as relações de risco observadas no estudo – por exemplo, que a hipertensão é um fator de risco para doença coronariana – podem ser generalizadas para todos os residentes de Framingham.

Então, o investigador deve considerar a validade de generalizar o achado de que a hipertensão é um fator de risco para doença coronariana nos adultos de Framingham para populações-alvo de outras partes do mundo. Essa inferência é mais subjetiva. A cidade de Framingham foi selecionada não a partir de uma estratégia científica de amostragem, e sim por ser uma típica cidade com população branca de classe média dos Estados Unidos e por ser conveniente para os investigadores. A validade de generalizar os resultados para outras populações baseia-se no conceito de que, em geral, estudos analíticos e ensaios clínicos sobre relações biológicas produzem resultados mais amplamente generalizáveis do que estudos descritivos sobre distribuições de características. Portanto, a força da associação entre hipertensão e doença coronariana é similar nos residentes brancos de Framingham do que em populações negras norte-americanas de baixa renda; no entanto, a prevalência de hipertensão é muito maior nesse segundo grupo.

Passos no delineamento de um protocolo para obtenção dos sujeitos do estudo

As inferências da Figura 3.2 são apresentadas da direita para a esquerda, sendo essa a sequência usada para interpretar os achados de um estudo já finalizado. No entanto, o investigador que está planejando um estudo inverte essa sequência, começando do lado esquerdo da figura (Figura 3.3). Ele começa especificando características clínicas e demográficas da

população-alvo adequadas à questão de pesquisa. Então utiliza **critérios geográficos** e **temporais** para especificar a escolha de uma amostra de estudo que seja ao mesmo tempo representativa e prática.

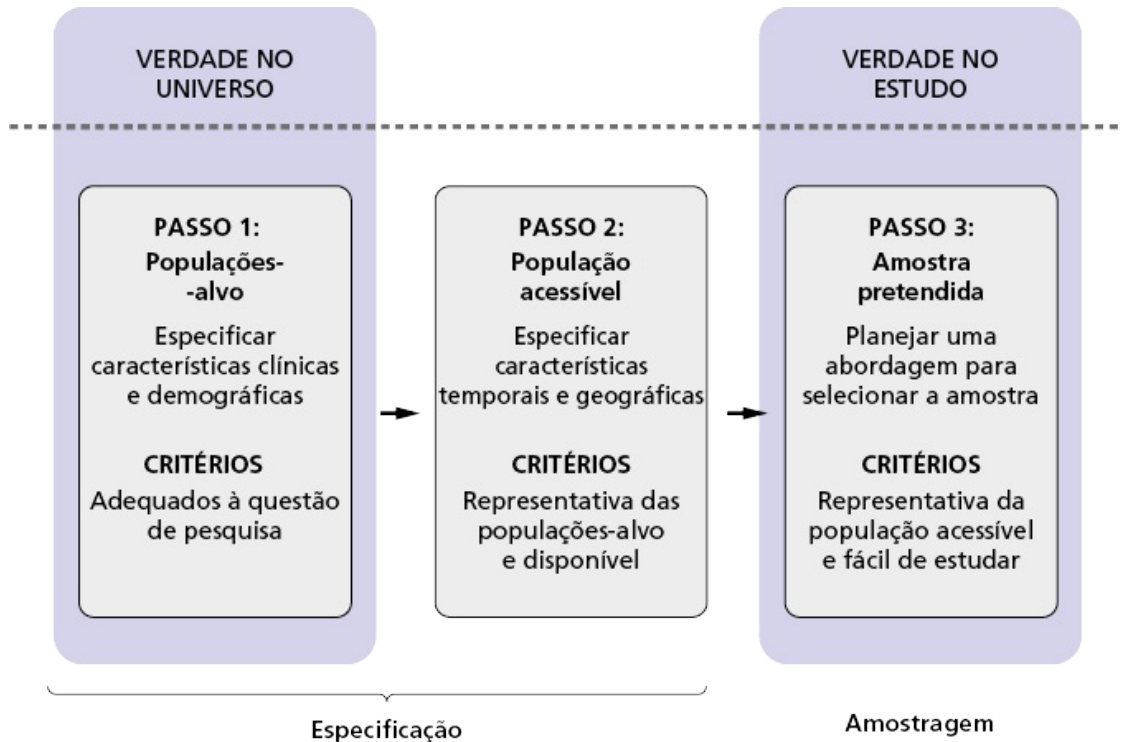



FIGURA 3.3 Passos no delineamento do protocolo para selecionar os sujeitos do estudo.

CRITÉRIOS DE SELEÇÃO

Se um investigador quer estudar a eficácia da suplementação de doses baixas de testosterona *versus* placebo para aumentar a libido em mulheres na menopausa, ele pode começar estabelecendo critérios de seleção que definem a população a ser estudada.

Estabelecendo critérios de seleção

Os **critérios de inclusão** definem as características principais da população-alvo relacionadas à questão de pesquisa (Tabela 3.1). A idade, na maioria das vezes, é um fator crucial e, nesse estudo, o investigador poderia escolher estudar mulheres na sexta década de vida, especulando que, nesse grupo etário, a relação entre risco e benefício desse medicamento pode ser a ideal; em outro estudo, poder-se-ia escolher

selecionar mulheres mais velhas. A opção também poderia ser por incorporar no estudo mulheres afro-americanas, hispânicas e asiáticas, em um esforço para aumentar a capacidade de generalização. Essa é geralmente uma boa ideia, mas é importante perceber que o aumento na capacidade de generalização é ilusório na ausência de outras evidências que sugiram que os efeitos diferem de acordo com a etnia. Nesse caso, o investigador precisaria de um número suficiente de mulheres de cada grupo étnico para testar estatisticamente para a presença de **modificação de efeito** (efeito em um grupo étnico diferente daquele em outros grupos étnicos, também denominado "interação"; ver Capítulo 9). Para isso, o número de sujeitos necessários geralmente é grande, e a maioria dos estudos não apresentam poder estatístico suficiente para detectar modificação de efeito.

TABELA 3.1 Delineando critérios de seleção para um ensaio clínico sobre testosterona em dose baixa versus placebo para aumentar a libido em mulheres na menopausa

	CARACTERÍSTICAS	EXEMPLO
Critérios de inclusão (seja específico)	Especificar populações relevantes à questão de pesquisa e eficientes para o estudo	
	Características demográficas	Mulheres com 50 a 59 anos de idade
	Características clínicas	Estar em bom estado de saúde Ter parceiro sexual Estar preocupada sobre diminuição da libido
	Características demográficas (administrativas)	Pacientes de um ambulatório do hospital do investigador
	Características temporais	Entre 1º de janeiro e 31 de dezembro do ano especificado
Critérios de exclusão (seja parcimonioso)	Especificar os subconjuntos da população que não serão estudados devido a:	
	Alta probabilidade de serem perdidos no seguimento	Alcoolistas Sujeitos que planejam se mudar para fora do Estado
	Incapacidade de fornecer dados confiáveis	Sujeitos desorientados Barreiras linguísticas*
	Alto risco de efeitos colaterais	História de infarto do miocárdio ou de AVE

* Alternativas à exclusão daqueles com barreira linguística (quando esses subgrupos forem grandes e importantes para a questão de pesquisa) poderiam ser coletar dados não verbais ou usar entrevistadores e questionários bilíngues.

Ao elaborar critérios de inclusão relacionados a características geográficas e temporais da população acessível, o investigador deve pesar as vantagens e desvantagens entre alcançar os objetivos científicos e os de ordem prática. Pacientes no próprio hospital podem ser mais acessíveis e apresentar menor custo. No entanto, as peculiaridades da população ou do ambiente local podem interferir na generalização dos resultados para outras populações. Nessas e em outras decisões sobre critérios de inclusão, não há uma única abordagem que seja claramente certa ou errada; o importante é tomar decisões sensatas que possam ser usadas de maneira uniforme ao longo do estudo e que possam ser claramente descritas para outras pessoas que tomarão a decisão sobre a quem as conclusões publicadas se aplicam.

Especificar características clínicas para selecionar os sujeitos para o estudo muitas vezes envolve decisões difíceis, não apenas sobre que fatores são relevantes para a questão de pesquisa, mas também sobre como defini-los. Por exemplo, como um investigador colocaria em prática o critério de que os sujeitos devem estar em boas condições de saúde? Ele poderia decidir não incluir pacientes com qualquer doença autorrelatada, mas isso provavelmente excluiria um grande número de sujeitos adequados para a questão de pesquisa.

Uma decisão melhor seria excluir apenas aquelas pessoas com doenças que poderiam interferir com o seguimento, como câncer metastático. Esse seria um exemplo de "**critério de exclusão**", que indica indivíduos os quais atendem aos critérios de inclusão e que seriam adequados para o estudo se não fosse por características que poderiam interferir no sucesso do seguimento, na qualidade dos dados ou na aceitabilidade da randomização (Tabela 3.1). A dificuldade com o idioma em que a pesquisa será realizada, problemas psicológicos, alcoolismo e doenças graves são exemplos de critérios de exclusão. Os **ensaios clínicos** diferem dos estudos observacionais por serem mais propensos a ter exclusões determinadas pela preocupação com a segurança de uma intervenção em certos pacientes; por exemplo, o uso de medicamentos em mulheres grávidas (Capítulo 10). Uma boa regra geral que mantém as coisas simples e preserva o número de sujeitos potenciais é ter o *menor número*

possível de critérios de exclusão.

População de base clínica versus população de base comunitária

Se a questão de pesquisa envolve pacientes com uma doença, recrutar sujeitos hospitalizados ou ambulatoriais é fácil e de baixo custo, mas os fatores de seleção que determinam quem vai ao hospital ou ao ambulatório podem ter um efeito importante. Por exemplo, um ambulatório especializado em um centro de atendimento médico terciário atrai pacientes de regiões distantes com formas graves da doença, dando uma visão distorcida das características e do prognóstico da doença em relação a como são vistos na prática diária. Amostras baseadas em ambulatórios de atenção primária podem ser uma opção melhor.

Outra opção comum de amostragem é selecionar sujeitos da comunidade que representem uma população de pessoas saudáveis. Essas amostras são frequentemente recrutadas usando correspondências, e-mails e anúncios na internet, na televisão ou na mídia impressa; elas não são inteiramente representativas de uma população geral porque alguns tipos de pessoas têm maior probabilidade de se voluntariar ou de usar a internet ou e-mail do que outros. Verdadeiras amostras de “base populacional” são difíceis e caras de recrutar, mas de grande utilidade em saúde pública e na prática clínica comunitária. Um dos maiores e melhores exemplos é o National Health and Nutrition Examination Survey (NHANES), que usou uma amostra **representativa** de todos os residentes dos Estados Unidos.

O tamanho e a diversidade de uma amostra podem ser ampliados por meio da colaboração com pesquisadores de outras cidades ou do uso de bancos de dados preexistentes, como o NHANES e os dados do Medicare. **Bancos de dados acessíveis eletronicamente** oriundos de agências de saúde pública, organizações de prestação de cuidados à saúde e companhias de seguros de saúde têm sido bastante usados em pesquisa clínica, tendem a ser mais representativos da população do país e a consumir menos tempo do que outras opções (Capítulo 13).

AMOSTRAGEM

A população de pessoas que atendem aos critérios de seleção costuma ser grande demais, sendo necessário selecionar uma **amostra** (subconjunto da

população) para estudo.

Amostragens não probabilísticas

Em pesquisa clínica, a amostra do estudo é frequentemente composta por indivíduos que atendem aos critérios de entrada e são de fácil acesso ao investigador. Por essa razão, recebe o nome de **amostragem de conveniência**. Esse tipo de amostragem tem vantagens óbvias em termos de custo e logística e é uma boa escolha para algumas questões de pesquisa.

Na **amostragem consecutiva,** pode-se minimizar o voluntarismo e outros tipos de viés de seleção, arrolando-se consecutivamente os sujeitos que atendem os critérios de entrada. Essa abordagem é especialmente desejável, por exemplo, quando resulta no arrolamento de toda a população acessível em um período de tempo que seja longo o suficiente para incluir variações sazonais ou outras mudanças temporais relevantes à questão de pesquisa.

A validade de se fazer inferências a partir de uma amostra depende do pressuposto de que, para responder à questão do estudo, ela represente adequadamente a população acessível. Em amostras de conveniência isso é meramente uma questão de julgamento.

Amostragens probabilísticas

Às vezes, em especial em estudos descritivos, é necessário fundamentar cientificamente a generalização dos achados da amostra do estudo para a população. As amostras probabilísticas, padrão-ouro na garantia da capacidade de generalização, usam um processo aleatório para assegurar que cada unidade da população tenha uma probabilidade especificada de seleção. É uma abordagem científica que fornece uma base rigorosa para estimar a fidelidade com que fenômenos observados na amostra representam aqueles da população e para computar significância estatística e intervalos de confiança. Há várias versões para essa abordagem.

- Na **amostragem aleatória simples**, enumeram-se (listam-se) todas as pessoas da população a partir da qual a amostra será sorteada e então seleciona-se aleatoriamente um subconjunto. O uso mais comum dessa

abordagem em pesquisa clínica é quando o investigador, diante de uma população maior que a necessária, seleciona um subconjunto representativo dela. Por exemplo, para sortear uma amostra aleatória dos pacientes submetidos a uma cirurgia de catarata no hospital, pode-se listar todos os pacientes das agendas do bloco cirúrgico no período do estudo e, então, usar uma tabela de números aleatórios para selecionar os indivíduos para a pesquisa (Apêndice 3).

- A **amostragem sistemática** se assemelha a uma amostragem aleatória simples no primeiro passo, quando se enumera a população; porém, difere no fato de que a seleção da amostra é feita por um processo periódico preordenado (p. ex., a abordagem de Framingham de selecionar as primeiras duas de cada três famílias de uma lista de famílias ordenadas por endereço). As amostras sistemáticas são suscetíveis a erros induzidos por periodicidades naturais da população e permitem ao investigador prever e, possivelmente, manipular quem entrará na amostra. Elas não oferecem vantagens logísticas em relação às amostras aleatórias simples e, em pesquisa clínica, raramente são a melhor opção.
- A **amostragem aleatória estratificada** inicia dividindo a população em subgrupos de acordo com características, como sexo ou raça, e selecionando uma amostra aleatória de cada um desses “estratos”. As subamostras estratificadas podem ser ponderadas para permitir seleção desproporcional de subgrupos menos comuns na população, mas que sejam de interesse especial ao investigador. Por exemplo, para investigar a incidência de toxemia gravídica, é possível estratificar a população de acordo com a raça e, então, sortear amostras de igual número de cada estrato. Grupos raciais menos comuns seriam, assim, hiper-representados, produzindo estimativas de incidência com precisões comparáveis para cada grupo.
- A **amostragem por conglomerados** é uma amostragem aleatória de agrupamentos naturais de indivíduos (conglomerados) na população. A amostragem por conglomerados é útil em populações bastante dispersas, nas quais se torna impraticável listar e amostrar todos os seus elementos. Considere, por exemplo, o problema de entrevistar pacientes

com câncer de pulmão sorteados aleatoriamente de um banco de dados que abrange os diagnósticos de alta hospitalar de todo o Estado. Seria mais eficiente selecionar uma amostra aleatória dos hospitais e, em seguida, os casos desses hospitais. Inquéritos comunitários costumam usar amostragem por conglomerados em dois estágios: sorteia-se uma amostra aleatória de quarteirões da cidade que estão enumerados em um mapa, então, uma equipe de campo visita os quarteirões sorteados, lista todos os endereços de cada um e seleciona uma subamostra dos endereços para estudo em um segundo processo aleatório. Uma desvantagem desse tipo de amostragem é que grupos que ocorrem naturalmente costumam ser mais homogêneos quanto às variáveis de interesse do que a população. Em cada quarteirão, por exemplo, o nível socioeconômico tende a ser semelhante. Isso significa que o tamanho de amostra efetivo (após ajuste para uniformidade intraconglomerado) será menor do que o número de sujeitos e que a análise estatística deve levar em conta os conglomerados.

Resumindo as opções de delineamento da amostragem

O uso de estatísticas descritivas e testes de significância estatística para inferir sobre a população a partir de observações na amostra parte do pressuposto de que uma amostragem probabilística foi empregada. No entanto, em pesquisa clínica, uma amostra aleatória da população-alvo raramente é viável. A amostragem de conveniência, preferentemente com seleção consecutiva, é uma abordagem prática apropriada para grande parte dos projetos de pesquisa clínica. No entanto, decidir se o delineamento proposto para a amostragem é satisfatório dependerá de um **julgamento** por parte do investigador: para a sua questão de pesquisa, as conclusões a partir das observações na amostra do estudo serão similares às que resultariam de uma verdadeira amostragem probabilística da população acessível? E, além disso, as conclusões serão apropriadas para a população-alvo?

RECRUTAMENTO

Metas para o recrutamento

Um fator importante para a escolha da população acessível e da estratégia

de amostragem é a viabilidade de recrutar os sujeitos para o estudo. Há duas metas básicas: (1) recrutar uma **amostra representativa** da população-alvo, minimizando a possibilidade de se chegar a uma resposta incorreta para a questão de pesquisa devido ao erro sistemático (viés); e (2) recrutar um **tamanho de amostra** suficiente para minimizar a possibilidade de se chegar a uma resposta incorreta devido ao erro aleatório (acaso).

Assegurando uma amostra representativa

A abordagem para recrutar uma amostra representativa começa na fase de delineamento, com decisões inteligentes sobre a escolha da população-alvo e da população acessível e com a definição da estratégia para realizar a amostragem. Termina na fase de implementação, quando são tomados cuidados para prevenir erros na aplicação dos critérios de entrada dos participantes e quando ocorre o aprimoramento das estratégias bem-sucedidas à medida que o estudo evolui.

Uma preocupação especial, principalmente em estudos descritivos, é o problema da **não resposta**.[1] A proporção dos sujeitos selecionados para o estudo que consentem em participar (taxa de resposta) influencia a validade da inferência de que a amostra arrolada representa a população. Os indivíduos difíceis de serem contatados e aqueles que se recusam a participar quando contatados tendem a ser diferentes dos que aceitam participar. O nível de não resposta que pode comprometer a capacidade de generalização do estudo depende da natureza da questão de pesquisa e dos motivos para a não resposta. Uma taxa de não resposta de 25%, embora satisfatória em muitos casos, pode distorcer gravemente a estimativa da prevalência de uma doença quando ela mesma é causa da não resposta. O grau em que o viés de não resposta influencia as conclusões de um estudo descritivo pode, às vezes, ser estimado durante o estudo, por meio da coleta de informações adicionais sobre uma subamostra dos não respondentes.

A melhor forma de lidar com o viés de não resposta, no entanto, é minimizá-lo desde o início. Para reduzir o número de indivíduos sorteados que não podem ser contatados, pode-se planejar uma série de tentativas repetidas de contato usando métodos alternativos (correio, e-mail, telefone, visita domiciliar). Entre as pessoas que são contatadas, a recusa

em participar pode ser minimizada das seguintes formas: melhorando a eficiência e a forma de despertar interesse sobre o estudo, escolhendo um delineamento que evite testes invasivos e que causem desconforto, usando folhetos explicativos e discussões individuais para aliviar a ansiedade e o desconforto, fornecendo incentivos como reembolso dos custos de transporte e resultados dos exames realizados e contornando barreiras linguísticas por meio de entrevistadores bilíngues e questionários traduzidos.

Recrutando um número suficiente de sujeitos

Taxa de recrutamento baixa é um dos problemas mais comuns na pesquisa clínica. Assim, ao planejar um estudo, é melhor pressupor que o número de sujeitos que irão atender os critérios de entrada e concordarão em participar será mais baixo, ou talvez muito mais baixo, que o número projetado no início. As abordagens para esse problema são estimar empiricamente com um pré-teste a magnitude do problema de recrutamento; planejar o estudo com uma população acessível que seja maior do que o que se considera necessário; e elaborar planos de contingência para o caso de serem necessários novos sujeitos. Enquanto o recrutamento está em andamento, é importante monitorar de perto o progresso das metas de recrutamento e sistematizar as razões pelas quais elas não tenham sido atingidas. Ao compreender por que motivo potenciais sujeitos são perdidos nos vários estágios do estudo, é possível chegar a estratégias para reduzir essas perdas.

Às vezes, o recrutamento envolve a seleção de sujeitos conhecidos da equipe de pesquisa (p. ex., um estudo sobre um novo tratamento em pacientes que frequentam a clínica do investigador). Nesse caso, a preocupação principal é dar a oportunidade de participação no estudo de forma adequada, deixando claras as reais vantagens e desvantagens. Ao discutir a participação, o investigador deverá reconhecer, em seus conselhos ao paciente, o dilema ético causado pelo conflito entre ser médico desse paciente e, ao mesmo tempo, investigador da pesquisa (Capítulo 14).

Muitas vezes, o recrutamento envolve o contato com populações desconhecidas aos membros da equipe de pesquisa. É fundamental, nesses casos, que ao menos um membro da equipe de pesquisa tenha experiência

prévia em abordagens para contatar esses potenciais sujeitos. Essas abordagens incluem o rastreamento no ambiente de trabalho ou em locais públicos como *shopping centers*; o envio de grande número de correspondências a endereços obtidos em listas como as de motoristas habilitados; o anúncio na internet; a solicitação a médicos conhecidos para que encaminhem pacientes; a revisão retrospectiva de prontuários; e o exame de listas de pacientes atendidos em contexto ambulatorial ou hospitalar. Algumas dessas abordagens, especialmente as duas últimas, envolvem preocupações com a invasão de privacidade que devem ser revisadas pelo Comitê de Ética em Pesquisa (CEP).

Na preparação para o recrutamento, pode ser útil obter o apoio de organizações importantes. Por exemplo, o investigador pode se reunir com o administrador do hospital, para discutir uma amostragem com base clínica, e com líderes comunitários e representantes de sociedades médicas e dos serviços municipais de saúde, para planejar uma operação de rastreamento na comunidade ou para enviar correspondências a médicos. Podem-se incluir cartas de apoio como anexos nas solicitações para financiamento. Para estudos de grande porte, pode ser útil criar um clima favorável na comunidade, por meio de palestras públicas ou anúncios em rádio, TV, jornais, cartazes, internet e malas diretas.

RESUMO

1. A maioria das pesquisas clínicas baseia-se, em termos filosóficos e práticos, no uso de uma **amostra** que representa uma **população**.
2. A vantagem de selecionar uma amostra é a **eficiência**. Permite ao investigador inferir sobre uma população grande, examinando apenas uma amostra a um custo relativamente pequeno em termos de tempo e esforço. A desvantagem é a possibilidade de **erro** que introduz. Se a amostra não for suficientemente representativa para a questão de pesquisa, os achados poderão não ser **generalizáveis** para a população-alvo, e, se ela não for suficientemente grande, os achados poderão não minimizar o suficiente o papel do **acaso**.
3. Ao planejar a amostra, o primeiro passo é conceitualizar a **população-alvo**. Isso significa formular um conjunto específico de **critérios de inclusão** que estabeleçam as características demográficas e clínicas

dos sujeitos adequados à questão de pesquisa.

4. O passo seguinte é selecionar uma **população acessível** adequada, que seja conveniente em termos geográficos e temporais, e definir um conjunto parcimonioso de **critérios de exclusão** que eliminem sujeitos cuja inclusão poderia ferir princípios éticos ou ser inapropriada para o estudo.
5. O próximo passo é delinear uma estratégia para a **amostragem** da população. Uma **amostra de conveniência** pode ser adequada, sobretudo para estudos preliminares sobre algumas questões, e uma **amostra consecutiva** frequentemente é uma boa opção. A **amostragem aleatória simples** pode ser usada para reduzir o tamanho da amostra, quando necessário. **Outras amostras probabilísticas** (**estratificada** e **por conglomerados**) podem ser úteis em determinadas situações.
6. Por fim, o investigador deve desenvolver e implementar estratégias que permitam o **recrutamento** de uma amostra de sujeitos que seja suficientemente **representativa** da população-alvo para controlar para fontes de erro sistemático e **suficientemente grande** para controlar para fontes de erro aleatório.

Apêndice 3

Esta tabela fornece uma forma simples e que dispensa auxílio de computador para selecionar uma amostra aleatória de 10% a partir de uma tabela de números aleatórios. Comece enumerando (listando e numerando) cada pessoa da população a ser amostrada. Então elabore uma regra para obter uma série de números apropriada: por exemplo, se a lista tiver 741 elementos (e você os numerou de 1 a 741), uma possível regra é descer verticalmente em cada coluna usando os três primeiros dígitos de cada número (começando em cima, no canto esquerdo, os números são 104, 223, etc.) e selecionar os primeiros 74 números distintos entre 1 e 741. Por fim, escolha um ponto de partida por um processo arbitrário (fechar os olhos e colocar o lápis em algum número da

tabela é uma forma de fazer isso) e comece a aplicar a regra. A abordagem moderna, que utiliza uma série computadorizada de números aleatórios, funciona basicamente da mesma forma.

TABELA 3.2 Selecionando uma amostra aleatória a partir de uma tabela de números aleatórios

10480	15011	01536	81647	91646	02011
22368	46573	25595	85393	30995	89198
24130	48390	22527	97265	78393	64809
42167	93093	06243	61680	07856	16376
37570	33997	81837	16656	06121	91782
77921	06907	11008	42751	27756	53498
99562	72905	56420	69994	98872	31016
96301	91977	05463	07972	18876	20922
89572	14342	63661	10281	17453	18103
85475	36857	53342	53998	53060	59533
28918	79578	88231	33276	70997	79936
63553	40961	48235	03427	49626	69445
09429	93969	52636	92737	88974	33488
10365	61129	87529	85689	48237	52267
07119	97336	71048	08178	77233	13916
51085	12765	51821	51259	77452	16308
02368	21382	52404	60268	89368	19885
01011	54092	33362	94904	31273	04146
52162	53916	46369	58569	23216	14513
07056	97628	33787	09998	42698	06691
48663	91245	85828	14346	09172	30163
54164	58492	22421	74103	47070	25306
32639	32363	05597	24200	38005	13363
29334	27001	87637	87308	58731	00256
02488	33062	28834	07351	19731	92420
81525	72295	04839	96423	24878	82651
29676	20591	68086	26432	46901	20949
00742	57392	39064	66432	84673	40027
05366	04213	25669	26422	44407	44048
91921	26418	64117	94305	26766	25940

REFERÊNCIA

1. www.framinghamheartstudy.org/about/background.html, acessado em 23/07/2012.

---

[1] A preocupação com a não resposta durante o recrutamento de sujeitos para um estudo (tema deste capítulo) ocorre primariamente em estudos descritivos cujo objetivo principal é estimar as distribuições de variáveis em determinadas populações. A não resposta durante o seguimento é um problema importante em qualquer estudo que acompanha uma coorte ao longo do tempo, especialmente em ensaios clínicos sobre intervenções que podem alterar a taxa de resposta (Capítulo 10).

CAPÍTULO 4

Planejando as aferições: precisão, acurácia e validade

Stephen B. Hulley, Thomas B. Newman e Steven R. Cummings

As aferições descrevem fenômenos em termos que podem ser analisados estatisticamente. Por sua vez, a validade de um estudo depende da capacidade que as variáveis delineadas têm de representar os fenômenos de interesse (Figura 4.1). Por exemplo, quão bem um glicosímetro portátil mede a glicemia, ou um questionário sobre insônia mede a quantidade e a qualidade do sono?

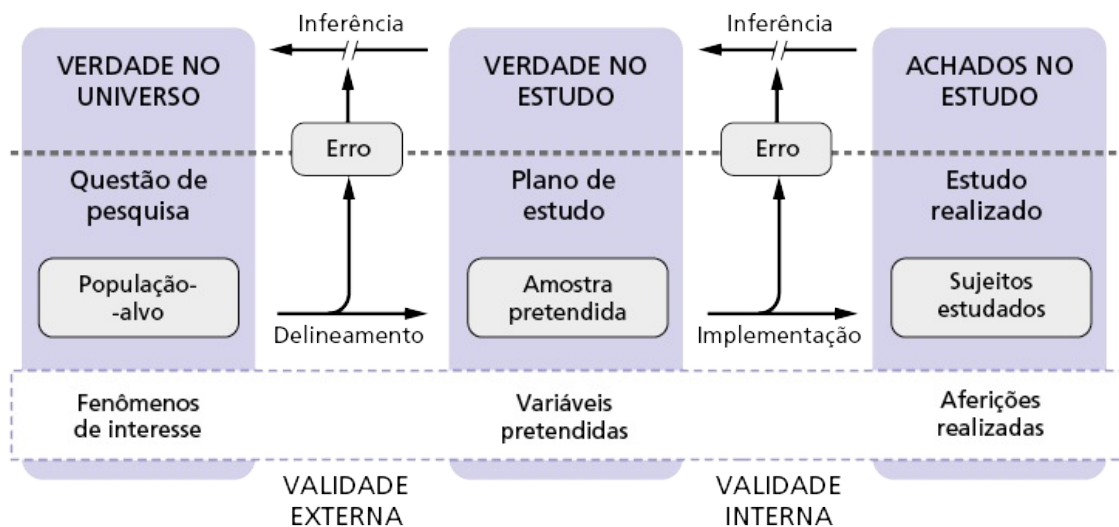



FIGURA 4.1 Delineando aferições que representem os fenômenos de interesse.

Este capítulo inicia tratando de como a escolha da **escala de medida** influencia o seu conteúdo informativo. Então abordamos a meta central, que é minimizar o erro de aferição. Isso envolve como delinear medidas que são relativamente **precisas** (livres de erro aleatório) e **acuradas**

(livres de erro sistemático), melhorando, assim, a adequação das inferências causais dessas aferições para os fenômenos de interesse. Em seguida, abordamos o conceito de **validade**, um parente qualitativo da acurácia. Por fim, concluímos com algumas considerações sobre aferições na pesquisa clínica e translacional, enfatizando as vantagens de armazenar espécimes para aferições posteriores.

ESCALAS DE MEDIDA

A Tabela 4.1 apresenta uma classificação simplificada das escalas de medida e o seu conteúdo informativo. Essa classificação é importante porque certos tipos de variáveis produzem **estatísticas mais informativas** que outros, o que aumenta o poder estatístico ou reduz as exigências de tamanho de amostra, além de permitir revelar padrões mais detalhados das distribuições.

TABELA 4.1 Escalas de medida

TIPO DE MEDIDA	CARACTERÍSTICAS DA VARIÁVEL	EXEMPLO	ESTATÍSTICAS DESCRITIVAS	PODER ESTATÍSTICO
		Categórica		
Dicotômica	Duas categorias	Estado vital (vivo ou morto)	Contagens, proporções	Baixo
Nominal	Categorias não ordenadas	Raça, tipo sanguíneo	Contagens, proporções	Baixo
Ordinal	Categorias ordenadas com intervalos não quantificáveis	Grau de dor, classe social	Contagens, proporções, medianas	Intermediário
		Numérica		
Contínua ou discreta*	Espectro ordenado com intervalos quantificáveis	Peso, número de cigarros/dia	Contagens, proporções, medianas, médias, desvios-padrões	Elevado

* As variáveis contínuas têm um número infinito de valores (p. ex., peso), ao passo que as variáveis discretas são mais limitadas (p. ex., número de cigarros/dia). Quando as variáveis discretas têm um número elevado de valores possíveis, assemelham-se a variáveis contínuas em termos práticos de poder e análise estatística.

Variáveis numéricas: contínuas e discretas

As **variáveis numéricas** podem ser quantificadas por meio de um número que expressa “quanto” ou “quantos”. As **variáveis contínuas** expressam “quanto” em uma escala infinita; o número de valores possíveis para o *peso corporal*, por exemplo, é apenas limitado pela sensibilidade do

equipamento usado para medi-lo. Por isso, diz-se que variáveis contínuas são muito informativas. As **variáveis numéricas discretas** expressam "quantos" em uma escala com unidades fixas, geralmente números inteiros, como o número de vezes em que uma mulher ficou grávida. Quando as variáveis discretas têm um número considerável de valores possíveis, elas se assemelham às variáveis contínuas nas análises estatísticas e podem ser consideradas equivalentes para fins de delineamento das aferições.

Variáveis categóricas: dicotômicas, nominais e ordinais

Fenômenos que não são facilmente quantificáveis podem ser aferidos por meio da classificação em categorias. **Variáveis categóricas** com dois valores possíveis (p. ex., *morto* ou *vivo*) são denominadas **dicotômicas**. As com mais de duas categorias (policotômicas) podem ser caracterizadas de acordo com o tipo de informação nelas contida. As **variáveis nominais** apresentam categorias não ordenadas, por exemplo, o sangue *tipo O* não é mais nem menos que o sangue *tipo B*, essas variáveis têm um caráter absoluto e qualitativo que as torna simples de medir. As categorias das **variáveis ordinais** apresentam uma ordem, como *dor forte, moderada ou leve*. Essa informação adicional é uma vantagem em relação às variáveis nominais, no entanto, por não especificarem uma diferença numérica ou uniforme entre as categorias, o seu conteúdo informativo é menor que o das variáveis numéricas discretas ou contínuas.

Escolhendo uma escala de medida

Uma boa regra geral é, sempre que for possível escolher, **preferir as variáveis contínuas em relação às categóricas**, pois a informação adicional nelas contida aumenta a eficiência estatística. Por exemplo, em um estudo que compara os efeitos anti-hipertensivos de diferentes tratamentos, medir a *pressão arterial em milímetros de mercúrio* permite ao investigador observar a magnitude da mudança em cada sujeito; medi-la como variável dicotômica (*hipertenso* vs. *normotenso*) limita o escopo dessa avaliação. Como as variáveis contínuas são muito mais informativas, elas possibilitam ao estudo um maior poder estatístico e/ou um menor tamanho de amostra (Capítulo 6).

As variáveis contínuas também permitem maior flexibilidade do que as

categóricas para ajustar os dados à natureza da variável ou ao formato da associação, especialmente quando a relação apresentar um padrão complexo. Por exemplo, em um estudo sobre a relação entre vitamina D e vários tipos de câncer, seria necessário medir a vitamina D como variável contínua para poder detectar um possível **padrão em forma de U**, ou seja, observar uma mortalidade maior em sujeitos com níveis baixos ou elevados de vitamina D do que nos indivíduos com níveis intermediários (1). Da mesma forma, em um estudo sobre os preditores do baixo peso ao nascer, deve-se registrar o peso ao nascer propriamente dito e não o fato de o peso estar acima ou abaixo do **ponto de corte** convencional de 2.500 g. Isso deixa a opção analítica em aberto: mudar o ponto de corte da definição de baixo peso ao nascer ou desenvolver uma escala ordinal com várias categorias de peso ao nascer (p. ex., > 2.500 g, 2.000-2.499 g, 1.500-1.999 g e < 1.500 g).

Da forma semelhante, quando há a possibilidade de definir o número de categorias de resposta em uma escala ordinal, como em uma questão sobre preferências alimentares, recomenda-se a adoção de uma meia dúzia de categorias que variam de "detesta" a "gosta muito". Os resultados podem, depois, ser reduzidos a uma dicotomia (*não gosta* e *gosta*), mas não vice-versa.

Muitas características, especialmente sintomas como dor ou aspectos relacionados ao estilo de vida, são difíceis de descrever em categorias ou números. No entanto, esses fenômenos são importantes para decisões diagnósticas e terapêuticas, e tentar medi-los é uma parte fundamental da abordagem científica de descrição e análise. Isso é ilustrado pelo Short Form (SF) –36, um questionário padronizado para avaliar a **qualidade de vida** que produz escores numéricos discretos (2). Os processos de classificação e aferição, quando feitos corretamente, podem tornar mais objetivos os nossos conhecimentos sobre o assunto, reduzir vieses e fornecer uma base de comunicação para a pesquisa.

PRECISÃO

Uma medida de alta **precisão** é aquela que é reprodutível, isto é, cujos valores são semelhantes em cada aferição. Uma balança pode medir o peso corporal com bastante precisão, mas uma entrevista sobre qualidade

de vida tem mais chances de produzir resultados que variam de acordo com o observador ou com a ocasião. A precisão tem uma influência importante no poder estatístico de um estudo. Quanto mais precisa for uma medida, maior o poder que um determinado tamanho de amostra tem para estimar os valores médios e testar hipóteses (Capítulo 6).

A precisão (também denominada reprodutibilidade, confiabilidade e consistência) é afetada pelo **erro aleatório** (variabilidade devida ao acaso): quanto maior o erro, menor é a precisão da aferição. Há três principais fontes de erro nas aferições:

- A **variabilidade do observador** é causada pelo observador e inclui aspectos como a escolha de palavras em uma entrevista e a habilidade no manuseio de um instrumento mecânico.
- A **variabilidade do instrumento** é causada pelo instrumento e inclui mudanças em fatores ambientais (p. ex., temperatura), desgaste de componentes mecânicos, diferenças entre os lotes dos reagentes, etc.
- A **variabilidade do sujeito** deve-se à variabilidade biológica intrínseca aos sujeitos da pesquisa e não está relacionada às variáveis em estudo. Inclui a variabilidade devido ao momento do dia em que é feita a aferição ou ao tempo transcorrido desde a última refeição ou medicação.

Avaliando a precisão

A precisão é avaliada como a **reprodutibilidade** de aferições repetidas, seja comparando aferições feitas pela mesma pessoa (reprodutibilidade intraobservador) ou por pessoas diferentes (reprodutibilidade interobservador). Da mesma forma, a reprodutibilidade também pode ser avaliada como inter e intrainstrumento. A reprodutibilidade de variáveis contínuas pode ser expressa como o **desvio-padrão intrassujeito** ou como o **coeficiente de variação** (desvio-padrão intrassujeito dividido pela média).[1] Para variáveis categóricas, são usados o **percentual de concordância**, o coeficiente de correlação intraclasse e a estatística **kapa** (3-5).

Estratégias para melhorar a precisão

Há cinco formas de minimizar o erro aleatório e aumentar a precisão das aferições (Tabela 4.2):

TABELA 4.2 Estratégias para reduzir o erro aleatório de forma a aumentar a precisão, com ilustrações de um estudo sobre tratamento anti-hipertensivo

ESTRATÉGIA PARA REDUZIR O ERRO ALEATÓRIO	FONTE DE ERRO ALEATÓRIO	EXEMPLO DE ERRO ALEATÓRIO	EXEMPLO DE ESTRATÉGIA PARA PREVENIR O ERRO
1. Padronização dos métodos de aferição em um manual de operações	Observador	Variação na aferição da pressão arterial (PA) causada pela variação na taxa de deflação do manguito (muitas vezes rápida demais)	Especificar que o manguito deve ser esvaziado a uma taxa de 2 mmHg/s
	Sujeito	Variação na PA decorrente da variação do tempo em que a pessoa está sentada em silêncio	Especificar que o sujeito deve sentar em uma sala silenciosa durante 5 minutos antes da aferição da PA
2. Treinamento e certificação do observador	Observador	Variação na PA em função da variação nas técnicas usadas pelo observador	Treinar o observador em técnicas-padrão
3. Otimização dos instrumentos	Instrumento ou observador	Variação na PA devido a um esfigmomanômetro malfuncionante	Adquirir um novo esfigmomanômetro de qualidade superior
4. Automatização do instrumento	Observador	Variação na PA decorrente da variação na técnica empregada pelo observador	Usar equipamento automático para medição da PA
	Sujeito	Variação na PA causada pela variação da reação emocional do sujeito ao observador	Usar equipamento automático para medição da PA
5. Repetição da aferição	Observador, sujeito e instrumento	Todas as aferições e todas as fontes de variação	Usar a média de duas ou mais medidas de PA

1. **Padronização dos métodos de aferição.** Todos os protocolos de estudo devem incluir instruções específicas para a realização de aferições (**definições operacionais**). Isso pode incluir instruções por escrito sobre como preparar o ambiente e o sujeito, como realizar e registrar a entrevista, como calibrar o instrumento e assim por diante (Apêndice 4). Esse conjunto de materiais, parte do **manual de operações**, é

fundamental para estudos grandes e complexos, mas altamente recomendado também para estudos menores. Mesmo quando houver apenas um observador, diretrizes específicas por escrito para cada aferição contribuem para que o desempenho seja uniforme ao longo do estudo e servem como base para descrever os métodos na hora de relatar os resultados.

2. **Treinamento e certificação dos observadores.** O treinamento melhora a consistência das técnicas de aferição, especialmente quando vários observadores estão envolvidos. É importante testar formalmente o domínio das técnicas especificadas no manual de operações e certificar-se de que os observadores alcançaram o nível necessário de desempenho (Capítulo 17).
3. **Otimização dos instrumentos.** Os instrumentos mecânicos e eletrônicos podem ser aperfeiçoados para diminuir a variabilidade. Da mesma forma, os questionários e as entrevistas podem ser redigidos de forma a ganhar clareza e evitar possíveis ambiguidades (Capítulo 15).
4. **Automatização de instrumentos.** Variações na forma como os observadores fazem as aferições podem ser eliminadas com dispositivos mecânicos automáticos e questionários de autorresposta.
5. **Repetição.** O efeito do erro aleatório de qualquer fonte é reduzido pela repetição das medições e uso da média de duas ou mais leituras. Essa estratégia aumenta muito a precisão. Suas principais limitações são o aumento do custo e as dificuldades de ordem prática envolvidas na repetição das aferições.

Para cada aferição no estudo, deve-se julgar a importância de se implementar essas estratégias. Sua adoção depende da importância da variável, da magnitude do potencial problema com a precisão e da factibilidade e custo da estratégia. De uma forma geral, as duas primeiras estratégias (padronização e treinamento) devem sempre ser usadas; a quinta (repetição), embora garanta o aumento de precisão, precisa ser factível e de custo acessível.

ACURÁCIA

A **acurácia** de uma variável é a sua capacidade de representar o valor verdadeiro.

A acurácia difere da precisão nos aspectos apresentados na Tabela 4.3. É importante ressaltar que acurácia e precisão não estão necessariamente relacionadas. Por exemplo, se o colesterol sérico fosse medido repetidamente usando padrões que haviam sido inadvertidamente diluídos duas vezes, faltaria acurácia ao resultado, mas, ainda assim, ele seria preciso (consistentemente errado por um fator de dois). Esse conceito é ilustrado na Figura 4.2. No entanto, acurácia e precisão normalmente andam juntas, e muitas das estratégias para aumentar a precisão também melhoram a acurácia.

TABELA 4.3 Precisão e acurácia da aferição

	PRECISÃO	ACURÁCIA
Definição	Grau em que uma variável tem valores semelhantes quando medida várias vezes	Grau em que uma variável se aproxim[a] do valor verdadeiro
Melhor forma de avaliar	Comparação entre medidas repetidas	Comparação com um "padrão-ouro"
Importância para o estudo	Aumento do poder estatístico para detectar os efeitos esperados	Aumento da validade das conclusões
Ameaçada por	Erro aleatório (acaso) causado pelo Observador Sujeito Instrumento	Erro sistemático (viés) causado pelo Observador Sujeito Instrumento

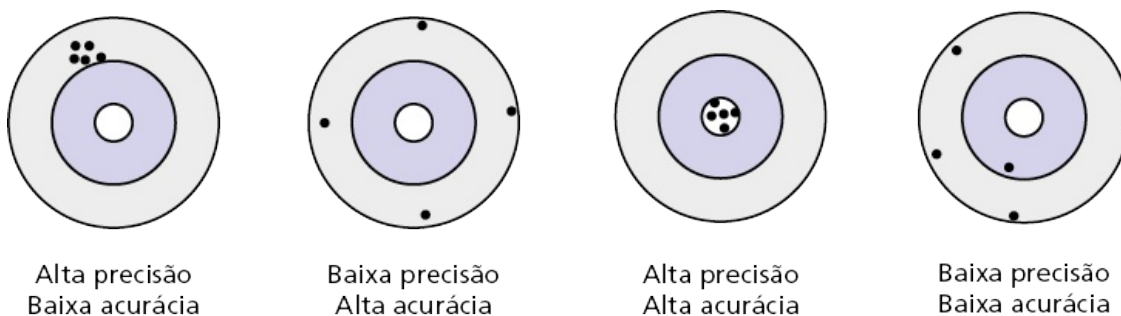



FIGURA 4.2 Diferença entre precisão e acurácia.

A acurácia é função do **erro sistemático:** quanto maior o erro, menor a acurácia da variável. As três principais classes de erro de aferição apontadas, na seção referente à precisão, têm seus equivalentes na acurácia apresentados a seguir:

- **Viés do observador** é uma distorção, consciente ou inconsciente, na

percepção ou no relato da medida pelo observador. Pode representar erros sistemáticos na forma de manuseio de um instrumento, como a tendência a arredondar para menos as medidas da PA, ou no uso de perguntas que induzem o entrevistado a uma determinada resposta.

- **Viés de instrumento** pode resultar de defeito em um instrumento mecânico. Por exemplo, uma balança que não foi calibrada recentemente pode apresentar valores mais baixos e começar a produzir repetidamente leituras mais baixas de peso corporal.

- **Viés do sujeito** é uma distorção na aferição originada pelo sujeito do estudo, por exemplo, ao relatar um evento (viés do respondedor). Por exemplo, as pacientes com câncer de mama que acreditam que o álcool seja uma causa do câncer que elas desenvolveram podem relatar uma quantidade exagerada de ingesta de álcool.

A acurácia de uma medida é mais bem avaliada comparando-a, quando possível, com um **padrão-ouro** – uma aferição de referência realizada usando uma técnica que se acredita representar o valor verdadeiro da característica. A decisão sobre que método de aferição será designado como padrão-ouro muitas vezes é uma decisão difícil, sendo necessário apoiar-se em trabalhos prévios já realizados sobre o assunto.

O grau de acurácia pode ser expresso, para medidas em escala contínua, como a diferença média entre a medida investigada e o padrão-ouro nos sujeitos do estudo. Para medidas em escala dicotômica, a acurácia em comparação com um padrão-ouro pode ser descrita em termos de sensibilidade e especificidade (veja Capítulo 12). Para medidas em escala categórica com mais de duas opções de resposta, pode-se calcular o percentual de respostas corretas para cada categoria.

Estratégias para melhorar a acurácia

As principais abordagens para aumentar a acurácia incluem as primeiras quatro estratégias listadas para a precisão e três outras (Tabela 4.4):

TABELA 4.4 Estratégias para reduzir o erro sistemático de forma a aumentar a acurácia, com ilustrações de um estudo sobre o tratamento anti-hipertensivo

ESTRATÉGIA PARA REDUZIR O ERRO SISTEMÁTICO	FONTE DE ERRO SISTEMÁTICO	EXEMPLO DE ERRO SISTEMÁTICO	EXEMPLO DE ESTRATÉGIA PARA PREVENIR O ERRO

1. Padronização dos métodos de aferição em um manual de operações	Observador	Leituras consistentemente elevadas da pressão arterial diastólica (PAD) decorrentes do uso do ponto de abafamento dos sons	Especificar a definição operacional da PAD como o ponto em que os sons se tornam inaudíveis
	Sujeito	Leituras consistentemente elevadas devido a PA ter sido medida logo após o sujeito ter subido as escadas para chegar ao ambulatório	Especificar que o sujeito fique sentado em uma sala silenciosa durante cinco minutos antes da aferição
2. Treinamento e certificação do observador	Observador	Leituras de PA consistentemente elevadas por terem sido empregados procedimentos diferentes dos especificados no manual de operações	O treinador verifica a acurácia da leitura do observador com um estetoscópio duplo
3. Otimização do instrumento	Instrumento	Leituras consistentemente elevadas da PA com um manguito-padrão em sujeitos com braços muito largos	Usar um manguito extra grande em pacientes obesos
4. Automatização do instrumento	Observador	Tendência consciente ou inconsciente do observador de ler valores mais baixos da PA no grupo randomizado para o medicamento ativo	Usar equipamento automático para aferição da PA
	Sujeito	Aumento da PA pela proximidade de um técnico ou uma técnica atraente	Usar equipamento automático para aferição da PA
5. Realização de aferições não intrusivas	Sujeito	Tendência do sujeito de superestimar a adesão ao medicamento estudado	Medir níveis do medicamento na urina
6. Calibração do instrumento	Instrumento	Leituras consistentemente elevadas da PA pelo fato de o manômetro aneroide estar descalibrado	Calibrar mensalmente
7. Cegamento	Observador	Tendência consciente ou inconsciente do observador de ler valores mais baixos de PA no grupo que recebeu tratamento ativo	Usar placebo duplo-cego para ocultar a alocação dos grupos de estudo
	Sujeito	Tendência dos sujeitos de super-	Usar placebo duplo-cego

	relatar os efeitos colaterais quando sabem que estão tomando o medicamento ativo	para ocultar a alocação dos grupos de estudo

1. **Padronização dos métodos de aferição.**
2. **Treinamento e certificação dos observadores.**
3. **Otimização dos instrumentos.**
4. **Automatização de instrumentos.**
5. **Realização de aferições não intrusivas.** Às vezes é possível fazer aferições sem que os sujeitos envolvidos fiquem cientes delas, eliminando, assim, a possibilidade de os sujeitos conscientemente enviesarem a medida. Por exemplo, uma avaliação sobre o efeito de se disponibilizar produtos para higienização das mãos e um cartaz estimulando a higienização das mãos em um refeitório de um hospital utilizou observadores que se misturaram com os usuários do refeitório (6).
6. **Calibração do instrumento.** A acurácia de muitos instrumentos, especialmente os mecânicos ou elétricos, pode ser aumentada com a calibração periódica contra um padrão-ouro.
7. **Cegamento (mascaramento).** Essa estratégia clássica não garante a acurácia total das medidas, mas pode eliminar **vieses diferenciais** que afetem um grupo de estudo mais do que outro. Em um ensaio clínico duplo-cego, o observador e o sujeito não sabem se o paciente recebeu o remédio ou o placebo, assegurando graus equivalentes de acurácia na medida de desfecho dos dois grupos.

Como citado em relação à precisão, a ênfase a ser dada a cada uma dessas sete estratégias fica a cargo do investigador. Os aspectos que fazem parte dessa análise incluem o potencial impacto da inacurácia sobre as conclusões do estudo e a factibilidade e o custo da estratégia. As primeiras duas estratégias (padronização e treinamento) devem ser sempre usadas; a calibração é necessária para todo instrumento que pode variar com o tempo; o cegamento é essencial quando factível.

VALIDADE

A **validade** se assemelha à **acurácia**, mas gostamos de vê-la como adicionando uma dimensão qualitativa à avaliação de até que ponto uma

medida representa adequadamente os fenômenos de interesse. Por exemplo, medidas da creatinina e da cistatina C séricas, dois compostos excretados pelos rins, podem ser igualmente acuradas (isto é, distanciando-se em até 1% do valor verdadeiro), mas a cistatina C pode ter maior validade como medida da função renal porque os níveis da creatinina são também influenciados pela quantidade de massa muscular (7). Na Figura 4.2, podemos pensar na validade como descrevendo se o centro do alvo é onde realmente queremos mirar.

Muitas vezes não é possível avaliar a validade por meio de um padrão-ouro, especialmente no caso de fenômenos subjetivos e abstratos, como dor ou qualidade de vida. Cientistas sociais desenvolveram construtos qualitativos e quantitativos para avaliar a validade dessas medidas.

- **Validade de conteúdo** avalia a capacidade da aferição de representar todos os aspectos dos fenômenos sob estudo – por exemplo, incluindo questões sobre o funcionamento social, físico, emocional e intelectual para avaliar a qualidade de vida.
- **Validade aparente** (*face validity*) descreve se as aferições parecem razoáveis, como no caso da aferição da dor usando uma escala de 10 pontos ou da classe social por meio da renda familiar.
- **Validade de construto** refere-se à capacidade de uma aferição de se encaixar dentro da concepção teórica (construto) sobre o fenômeno em estudo; por exemplo, um teste de QI deveria distinguir entre pessoas que, segundo a teoria ou outras medidas, teriam níveis diferentes de inteligência.
- **Validade preditiva** refere-se à capacidade da medida de predizer a ocorrência futura de um desfecho, por exemplo, a capacidade de um questionário desenvolvido para avaliar depressão em predizer a perda do trabalho ou o risco de suicídio.
- **Validade de critério** é o grau em que a medida se correlaciona com medidas já existentes e bem aceitas.

A abordagem geral para medir fenômenos subjetivos e abstratos inicia com uma revisão da literatura e uma consulta a especialistas para localizar

um **instrumento** adequado (em geral, um questionário) que já tenha sido validado. Essa estratégia tem a vantagem adicional de tornar os resultados do novo estudo comparáveis a trabalhos anteriores na área, podendo simplificar e fortalecer o processo de solicitação de financiamento e publicação dos resultados. As suas desvantagens, no entanto, residem no fato de que a validação pode não ter sido feita da melhor forma e que um instrumento antigo desengavetado pode ser antiquado ou não apropriado à questão de pesquisa.

Se os instrumentos existentes não se adequarem às necessidades do estudo, o investigador poderá decidir desenvolver uma nova abordagem de medição e validá-la ele mesmo. Isso pode ser um desafio interessante e até mesmo levar a uma contribuição importante para a literatura, mas o processo geralmente leva tempo e é trabalhoso (Capítulo 15). Além disso, é preciso lembrar que o processo, em geral, é menos conclusivo que o conotado pela palavra “validação”.

OUTRAS CARACTERÍSTICAS DE ABORDAGENS DE AFERIÇÃO

As medidas devem ser suficientemente **sensíveis** para detectar diferenças que são importantes para o investigador. O grau de sensibilidade necessário depende da questão de pesquisa. Por exemplo, um estudo que avalia se um novo medicamento ajuda a parar de fumar pode usar uma medida de desfecho não muito sensível ao número exato de cigarros fumados por dia. Por outro lado, se a questão é o efeito da redução de nicotina nos cigarros sobre o número de cigarros fumados, o método de aferição precisaria ser sensível a mudanças de apenas alguns cigarros diários.

A medida ideal é **específica**, isto é, representa apenas a característica de interesse. O nível de monóxido de carbono no ar expirado é uma medida apenas moderadamente específica do hábito de fumar, pois tal medida também pode ser afetada pela exposição às emissões de automóveis, entre outras. A especificidade na avaliação do hábito de fumar pode ser aumentada acrescentando medidas (como autorrelato e nível de cotinina sérica) que não são afetadas pela poluição do ar.

As medidas devem ser **adequadas** aos objetivos do estudo. Por

exemplo, para estudar o efeito do estresse no infarto do miocárdio, antes de começar as definições operacionais das aferições, seria necessário decidir qual o tipo de estresse (psicológico ou físico, agudo ou crônico) que será investigado.

As medidas devem fornecer uma ampla **distribuição de respostas** na amostra do estudo. Uma medida do estado funcional tem utilidade máxima quando produz valores que variam de alto, em alguns sujeitos, a baixo, em outros.

Um dos principais motivos para realizar um pré-teste é garantir que as respostas reais não se concentrem em um extremo da faixa de respostas possíveis (Capítulo 17).

Sempre que possível, as aferições devem ser planejadas de modo que minimizem julgamentos subjetivos. A **objetividade** é alcançada reduzindo-se o envolvimento do observador e utilizando instrumentos automatizados. No entanto, um risco dessas estratégias é produzir uma visão em túnel, limitando o escopo das observações e a capacidade de descobrir fenômenos não antecipados. Uma forma de lidar com esse problema é incluir algumas questões abertas e adquirir dados subjetivos e qualitativos para complementar o conjunto principal de medidas objetivas e quantitativas.

Ao delinear um estudo, há uma tendência a ficar acrescentando itens que não são centrais à questão de pesquisa, mas que *poderiam* ser de interesse. É verdade que medições adicionais aumentam a probabilidade de encontrar achados interessantes, incluindo alguns que não haviam sido antecipados desde o início. Entretanto, é importante ter em mente o valor da **eficiência** e da **parcimônia**. O conjunto completo de aferições deve ser delineado para incluir dados relevantes a um custo acessível em termos financeiros e de tempo a ser despendido. Coletar dados em excesso é um erro comum que pode cansar os sujeitos, sobrecarregar a equipe que está fazendo as aferições e complicar o manejo e a análise estatística dos dados. Isso pode resultar em um estudo mais dispendioso e, paradoxalmente, não tão bem-sucedido na resposta às questões principais de pesquisa.

MEDIÇÕES EM MATERIAIS ARMAZENADOS

A pesquisa clínica envolve aferições em indivíduos que variam em um amplo espectro de domínios. Algumas aferições podem ser feitas apenas durante o contato com o sujeito do estudo, mas muitas podem ser feitas posteriormente, em **bancos de amostras biológicas** armazenadas para análise química ou genética, ou quando **imagens** de radiografia e de outros procedimentos são armazenadas eletronicamente (Tabela 4.5).

TABELA 4.5 Tipos comuns de aferições que podem ser feitas em materiais armazenados

TIPO DE MEDIDA	EXEMPLOS	BANCO PARA AFERIÇÃO POSTERIOR
História médica	Diagnósticos, medicamentos, cirurgias, sintomas, achados do exame físico	Prontuários eletrônicos ou em papel
Fatores psicossociais	Depressão, história familiar	Gravações de áudio e vídeo
Antropometria	Altura, peso, composição corporal	Fotografias
Medidas bioquímicas	Colesterol sérico, fibrinogênio plasmático	Soro, plasma, urina, espécimes histopatológicos
Testes genéticos/moleculares	Polimorfismos de nucleotídeo único	DNA
Imagem	Densidade óssea, cálcio coronariano	Raio X, tomografia computadorizada, ressonância magnética
Eletromecânica	Arritmia, cardiopatia congênita	Eletrocardiograma, ecocardiograma

Uma vantagem desse tipo de armazenamento é a oportunidade de reduzir o custo fazendo aferições apenas em indivíduos que desenvolvem a condição de interesse durante o estudo. O delineamento de caso-controle aninhado (Capítulo 8) é uma opção excelente para esse fim, especialmente se for possível fazer aferições cegas e pareadas no mesmo ensaio analítico, eliminando, assim, o componente interensaio do erro aleatório. Outra grande vantagem é permitir que avanços científicos que ocorrerem anos após o início do estudo possam levar a novas ideias e técnicas de aferição, que poderão, então, ser empregadas, financiadas por novos auxílios de pesquisa.

O interesse crescente na **pesquisa translacional** (Capítulo 2) se beneficia de técnicas novas de aferição que expandiram enormemente a abrangência da pesquisa clínica, por exemplo, nas áreas da **epidemiologia genética e molecular** (8, 9) e de **exames de imagem**. As aferições em amostras biológicas que contêm DNA (p. ex., saliva, sangue) trazem

informações novas sobre genótipos que podem contribuir para a ocorrência de uma doença ou modificar a resposta de um paciente ao tratamento. Medidas séricas podem ser usadas para estudar as causas ou consequências moleculares de uma doença, por exemplo, marcadores inflamatórios podem fornecer informações úteis sobre a fisiopatologia de muitas doenças. É importante consultar com especialistas para selecionar os tubos adequados para coleta e implementar as condições adequadas de armazenamento, de modo a preservar a qualidade das amostras e permitir que estejam disponíveis para o maior leque possível de usos subsequentes. Também é importante obter o consentimento informado dos participantes, informando sobre o escopo dos potenciais usos das amostras armazenadas.

RESUMO

1. As variáveis podem ser **numéricas** ou **categóricas**. As variáveis numéricas são **contínuas** (quantificadas em uma escala infinita) ou **discretas** (quantificadas em uma escala finita, como números inteiros); as variáveis categóricas são **nominais** (não ordenadas) ou **ordinais** (ordenadas), e aquelas com apenas duas categorias são denominadas **dicotômicas**.
2. **Variáveis mais informativas** permitem maior poder estatístico e/ou menor tamanho de amostra, de acordo com a seguinte **hierarquia**: variáveis contínuas > variáveis numéricas discretas > variáveis ordinais > variáveis nominais e dicotômicas.
3. A **precisão** de uma medida (reprodutibilidade de aferições repetidas) é outro determinante importante do poder estatístico e do tamanho da amostra. A precisão é reduzida pelo **erro aleatório** (acaso) a partir de três fontes de variabilidade: observador, sujeito e instrumento.
4. As estratégias para **aumentar a precisão** que devem ser parte de todo estudo são **definir** e **padronizar operacionalmente os métodos** em um **manual de operações**. Outras estratégias úteis são **treinar e certificar os observadores**, **otimizar e automatizar os instrumentos** e usar a média de **aferições repetidas**.
5. A **acurácia** de uma medida é o grau em que ela se aproxima de um padrão-ouro. A acurácia é reduzida por **erro sistemático** (**viés**),

resultante das mesmas três fontes: observador, sujeito e instrumento.

6. As **estratégias para aumentar a acurácia** incluem todas as listadas em relação à precisão, com exceção da repetição. Além disso, a acurácia é aumentada com **aferições não intrusivas**, **calibração** e, nas comparações entre grupos, por **cegamento**.
7. A **validade** é o grau em que uma medida representa os fenômenos que ela deveria medir. Ela é comumente usada para variáveis mais abstratas e subjetivas e é avaliada por meio da **validade de conteúdo**, **validade aparente** (*face validity*), **validade de construto**, **validade preditiva** e **validade de critério**.
8. As medidas devem ser **sensíveis**, **específicas**, **adequadas** às questões do estudo e **objetivas**, além de capazes de produzir uma **faixa ampla de valores**. Em suma, devem ser amplas, mas **parcimoniosas**, servindo à questão de pesquisa a um custo aceitável em termos de tempo e recursos financeiros.
9. Os investigadores devem considerar **estocar imagens e outros materiais** para medições posteriores que se beneficiem de novas tecnologias à medida que elas forem desenvolvidas, e da eficiência dos delineamentos do tipo caso-controle aninhado.

APÊNDICE 4

MANUAL DE OPERAÇÕES: DEFINIÇÃO DE UMA MEDIDA DE FORÇA DE PREENSÃO MANUAL

O manual de operações descreve o método para conduzir e registrar os resultados de todas as aferições feitas no estudo. Este exemplo foi retirado do manual de operações do nosso Estudo sobre Fraturas Osteoporóticas. Ele descreve o uso de um dinamômetro para medir a força de preensão manual. Para padronizar as instruções de examinador a examinador e de sujeito a sujeito, o protocolo inclui um roteiro de instruções que devem ser lidas, palavra por palavra, ao participante.

PROTOCOLO PARA MEDIR A FORÇA DE PREENSÃO MANUAL COM O DINAMÔMETRO

A força de preensão manual será medida em ambas as mãos. O tamanho da mão deve ser ajustado para que o participante segure o dinamômetro confortavelmente. Coloque o dinamômetro na mão direita com o indicador voltado para a palma da mão. O braço do participante deve ser fletido na altura do cotovelo a um ângulo de 90°, com o antebraço paralelo ao chão.

1. Demonstre o teste ao sujeito. Ao demonstrar, instrua o indivíduo usando a seguinte descrição: "Esse aparelho mede a força de seu braço e da parte superior de seu corpo. Vamos medir sua força de preensão em ambos os braços. Vou demonstrar como proceder. Dobre o cotovelo a um ângulo de 90°, com o antebraço paralelo ao chão. Não deixe o braço tocar a lateral de seu corpo. Baixe o aparelho lentamente e o aperte o mais forte que puder, enquanto eu conto até três. Uma vez que seu braço estiver completamente estendido, você pode soltar a mão".
2. Deixe o paciente treinar uma vez para cada braço, iniciando com o braço direito se for destro. Na segunda vez, registre o número de kg-força apontado pelo indicador, com a precisão de 0,5 kg.
3. Zere o indicador. Repita os procedimentos para o outro braço.

O braço não deve tocar o corpo. A preensão deve ser um aperto lento e sustentado, não um movimento brusco e explosivo.

REFERÊNCIAS

1. Michaelsson K, Baron JA, Snellman G, et al. Plasma vitamin Δ and mortality in older men: a community-based prospective cohort study. *Am J Clin Nutr* 2010;92:841–848.
2. Ware JE, Gandek β Jr. Overview of the SF-36 health survey and the International Quality of Life Assessment Project. *J Clin Epidemiol* 1998;51:903–912.
3. Bland JM, Altman DG. Measurement error and correlation coefficients. *BMJ* 1996;313:41–42; also, Measurement error proportional to the mean. *BMJ* 1996;313:106.
4. Newman TB, Kohn M. Evidence-based diagnosis. New York: Cambridge University Press, 2009.
5. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*

1960;20:37– 46.

6. Filion K, Kukanich KS, Chapman B, et al. Observation-based evaluation of hand hygiene practices and the effects of an intervention at a public hospital cafeteria. *Am J Infect Control* 2011;39:464–470.
7. Peralta CA, Shlipak MG, Judd S, et al. Detection of chronic kidney disease with creatinine, cystatin C, and urine albumin-to-creatinine ratio and association with progression to end-stage renal disease and mortality. *JAMA* 2011;305:1545–1552.
8. Guttmacher AE, Collins FS. Genomic medicine: a primer. *NEJM* 2002;347:1512–1520.
9. Healy DG. Case–control studies in the genomic era: a clinician's guide. *The Lancet Neurology* 2006; 5:701–707.

1Quando são feitas duas aferições de uma variável contínua para cada sujeito, pode ser tentador expressar sua concordância usando um **coeficiente de correlação**. Entretanto, uma vez que o coeficiente de correlação é muito sensível a valores extremos (*outliers*) (3, 4), uma abordagem melhor é o gráfico de Bland-Altman, no qual a diferença entre as duas aferições é plotada como uma função de sua média. Se o valor absoluto da diferença entre as aferições aumentar linearmente com a média, o coeficiente de variação é uma forma melhor de sumarizar a variabilidade do que o desvio-padrão intrassujeito.

CAPÍTULO 5

Preparando-se para estimar o tamanho de amostra: hipóteses e princípios básicos

Warren S. Browner, Thomas B. Newman e Stephen B. Hulley

Uma vez decidido o que e quem estudar e que delineamentÍo será usado, é preciso decidir quantos sujeitos deverão compor a amostra. Até mesmo o estudo mais rigorosamente executado poderá não responder à questão de pesquisa se o tamanho de amostra for insuficiente. Por outro lado, um estudo com amostra muito grande traz mais dificuldades e custos do que o necessário. A meta do planejamento do tamanho de amostra é estimar um **número adequado** de sujeitos para um dado delineamento de pesquisa.

Embora possam servir como guia útil, os cálculos de tamanho de amostra dão a falsa impressão de objetividade estatística. Sua acurácia depende inteiramente dos dados e das estimativas empregados, que muitas vezes são apenas adivinhações ou "chutes" informados. É útil pensar no **planejamento do tamanho de amostra** como uma forma matemática de fazer uma estimativa grosseira. Muitas vezes, isso revela que o delineamento do estudo não é factível ou que é preciso mudar as variáveis preditoras ou de desfecho. Portanto, o tamanho de amostra deve ser estimado em um estágio inicial do processo de delineamento do estudo, quando ainda é possível fazer mudanças maiores.

Antes de apresentarmos, no Capítulo 6, como calcular o tamanho de amostra para vários delineamentos comuns de pesquisa, vamos tratar aqui dos princípios que norteiam esses cálculos. Os leitores que tiverem dificuldade em compreender esses princípios ficarão aliviados ao descobrir que não é necessário dominá-los para poder planejar o tamanho

de amostra. Todavia, da mesma forma que uma receita faz mais sentido se o cozinheiro conhecer os ingredientes, o cálculo de tamanho de amostra é mais fácil se o investigador tiver familiaridade com os conceitos básicos. Mesmo se você pedir auxílio a um bioestatístico para calcular o tamanho de amostra para o seu estudo, entender como o processo funciona irá auxiliá-lo a participar mais ativamente na consideração dos pressupostos e estimativas envolvidos nesse cálculo.

HIPÓTESES

O processo começa pela reformulação de sua questão de pesquisa como uma **hipótese de pesquisa** que resume os elementos principais do estudo – a amostra e as variáveis preditora e de desfecho – de uma forma que estabelece a base para os testes de significância estatística. Por exemplo, suponha que sua questão de pesquisa seja sobre se as pessoas que jogam palavras cruzadas têm menor probabilidade de desenvolver demência. Sua hipótese de pesquisa precisaria especificar a amostra (p. ex., pessoas que moram em um lar de idosos e que têm função cognitiva normal), a variável preditora (jogar palavras cruzadas pelo menos uma vez por semana, em média) e a variável de desfecho (escore anormal em um teste padronizado sobre função cognitiva após dois anos de seguimento).

As hipóteses não são necessárias em estudos descritivos sobre a distribuição das características em uma população, como a prevalência de função cognitiva anormal em um lar de idosos. (Isso não significa, contudo, que não é necessário fazer uma estimativa do tamanho de amostra para um estudo descritivo; significa apenas que os métodos para fazê-lo, descritos no Capítulo 6, são diferentes.) As hipóteses são importantes, no entanto, em estudos que fazem testes estatísticos para comparar diferentes grupos, como se idosos que jogam palavras cruzadas regularmente têm menor probabilidade de desenvolver demência. Visto que a questão de pesquisa da maioria dos estudos observacionais e de todos os experimentais envolve comparações, ela exige a especificação de pelo menos uma hipótese. Se algum dos termos a seguir aparecer na questão de pesquisa, o estudo não será meramente descritivo, sendo, assim, necessário formular uma hipótese de pesquisa: maior que, menor que, mais provável que, relacionado a, associado a, comparado com,

semelhante a, correlacionado com, causa, ou leva a.

Características de uma boa hipótese de pesquisa

Uma boa hipótese deve basear-se em uma boa questão de pesquisa. Além disso, deve ser simples, específica e formulada *a priori*.

Simples *versus* complexa

Uma **hipótese simples** contém uma variável preditora e uma de desfecho:

Em pacientes com diabetes tipo 2, um estilo de vida sedentário está associado a um maior risco de desenvolver proteinúria.

Uma **hipótese complexa** contém mais de uma variável preditora:

Em pacientes com diabetes tipo 2, um estilo de vida sedentário e o consumo de álcool estão associados a um maior risco de desenvolver proteinúria.

Ou mais de uma variável de desfecho:

Em pacientes com diabetes tipo 2, o consumo de álcool está associado com risco aumentado de desenvolver proteinúria e neuropatia.

Hipóteses complexas como essas não são prontamente testáveis com um teste estatístico único, sendo mais fácil abordá-las na forma de duas ou mais hipóteses simples. No entanto, uma alternativa possível é usar uma variável preditora ou de desfecho combinada:

Em pacientes com diabetes tipo 2, o consumo de álcool está associado a um maior risco de desenvolver complicação microvascular – isto é, proteinúria, neuropatia ou retinopatia.

Nesse último exemplo, o investigador decidiu que o importante é se o participante apresenta alguma complicação microvascular, e não o tipo dessa complicação.

Específica *versus* vaga

Uma **hipótese específica** não deixa ambiguidade sobre os sujeitos e as variáveis ou sobre como o teste de significância estatística será aplicado. Ela inclui definições operacionais concisas que resumem a natureza e a fonte dos sujeitos e como as variáveis serão aferidas.

O uso prévio de antidepressivos tricíclicos por pelo menos seis semanas é mais comum em pacientes hospitalizados por infarto do miocárdio no Hospital de Longview do que nos controles hospitalizados por pneumonia.

Essa frase pode parecer longa, mas comunica a natureza do estudo de forma clara, minimizando a possibilidade de que algo um pouco diferente seja testado na hora de examinar os dados. Seria incorreto, por exemplo, substituir, durante a fase de análise, a variável preditora por uma outra forma de medi-la, como depressão autorrelatada, sem considerar o problema de testar hipóteses múltiplas (tópico que será discutido no final deste capítulo). Para manter a hipótese de pesquisa concisa, não é necessário enunciar todos esses elementos, que podem ser explicitados no plano de estudo. No entanto, esses detalhes devem sempre estar claros na concepção do investigador sobre o estudo e especificados no procotolo.

Às vezes, fica evidente na hipótese de pesquisa se a variável preditora e a de desfecho são dicotômicas, contínuas ou categóricas. Caso isso não fique claro, o tipo de variável deve ser especificado:

Em homens não obesos de 35 a 59 anos, participar em uma liga de boliche pelo menos uma vez por semana está associado com um maior risco de desenvolver obesidade (índice de massa corporal > 30 kg/m^2) durante um seguimento de 10 anos.

Novamente, se a hipótese de pesquisa ficar detalhada demais, as definições podem ser omitidas, contanto que sejam explicitadas em outro local.

Antes ou depois dos fatos

A hipótese deve ser formulada por escrito no início do estudo, o que leva o investigador a focar o esforço da pesquisa em seu objetivo central. Além disso, a pré-formulação de uma única hipótese cria uma base mais forte para a interpretação dos resultados do estudo do que várias hipóteses surgidas na inspeção dos dados. Hipóteses formuladas após o exame dos dados são uma forma de testar hipóteses múltiplas, que, com frequência, levam à interpretação exagerada da importância dos resultados.

Hipóteses nula e alternativa

Atenção: Se você não teve treino formal em estatística ou se esqueceu o que aprendeu, os parágrafos a seguir não farão muito sentido em uma primeira leitura. Tente revisar a terminologia, mesmo se ela parecer muito complicada ou estranha.

O processo inicia por uma reformulação da hipótese de pesquisa de modo a propor que não há diferença entre os grupos sob comparação. Essa reformulação, denominada **hipótese nula**, será a base formal para testar a significância estatística quando você for analisar os dados no final do estudo. Partindo do pressuposto de que não há associação na população, testes estatísticos podem ajudar a estimar a probabilidade de que uma eventual associação observada em um estudo pode se dever ao acaso.

Por exemplo, suponha que sua questão de pesquisa seja sobre se tomar água de torneira não filtrada está associado a um risco aumentado de desenvolver úlcera péptica (talvez devido a um maior risco de contaminação por *H. pylori*). Sua hipótese nula – a de que não há associação entre as variáveis preditora e de desfecho na população – seria:

> *Pessoas em Phnom Penh que tomam água de torneira não filtrada têm o mesmo risco de desenvolver úlcera péptica do que as que tomam água mineral.*

A proposição de que *há* associação ("Pessoas em Phnom Penh que tomam água de torneira têm risco maior de desenvolver úlcera péptica do que as que tomam água mineral.") é denominada **hipótese alternativa**. A hipótese alternativa não pode ser testada diretamente; o procedimento-padrão é aceitá-la se o teste de significância estatística rejeitar a hipótese nula (veja mais adiante).

Devemos também revisar agora outra terminologia confusa. A hipótese alternativa pode ser uni ou bilateral. A **hipótese alternativa unilateral** (também denominada unicaudal ou unidirecional) especifica a direção da associação entre as variáveis preditora e de desfecho. Um exemplo de hipótese unilateral é a de que tomar água de torneira aumenta o risco de úlcera péptica (em comparação com água mineral). A **hipótese alternativa bilateral** (bicaudal ou bidirecional) declara apenas que há associação, sem especificar em que direção. Por exemplo, "Tomar água

de torneira está associado com um risco diferente – aumentado ou diminuído – de desenvolver úlcera péptica do que tomar água mineral”.

As hipóteses unilaterais poderiam ser apropriadas em determinadas circunstâncias, como quando apenas uma direção para uma associação é clinicamente importante ou biologicamente significativa. Um exemplo desse caso é quando se testa se um novo medicamento para hipertensão tem maior probabilidade de causar erupções cutâneas do que o placebo, não interessando a possibilidade de o medicamento causar menos erupções cutâneas (mas esse poderia ser o caso se o medicamento tivesse propriedades anti-inflamatórias). Outro caso adequado ao uso de uma hipótese unilateral é quando há evidências fortes em estudos anteriores de que uma associação em uma das duas direções é improvável, como uma hipótese sobre se o fumo de cigarros afeta o risco de câncer cerebral. Nesse caso, a hipótese unilateral é justificada por evidências anteriores sobre a baixa probabilidade de o fumo diminuir a incidência de câncer cerebral, uma vez que o fumo aumenta o risco de vários tipos de câncer. No entanto, é importante atentar para o fato de que muitas hipóteses bem-embasadas se enfraquecem quando testadas por ensaios clínicos randomizados. Dois exemplos disso são as seguintes hipóteses *a priori*: a terapia com β-caroteno reduz o risco de câncer de pulmão, e o tratamento para redução de extrassístoles ventriculares reduz a morte súbita em pacientes com arritmias ventriculares. Nesses dois exemplos, resultados de ensaios clínicos randomizados bem-conduzidos revelaram um efeito estatisticamente significativo que teve direção oposta ao que os investigadores esperavam encontrar (1-3). De maneira geral, acreditamos que a maioria das hipóteses alternativas deveriam ser bilaterais.

No entanto, é importante ter em mente a diferença entre a **hipótese de pesquisa**, geralmente unilateral, e a hipótese alternativa, usada no planejamento do tamanho da amostra, que é quase sempre bilateral. Por exemplo, considere a questão de pesquisa sobre se o uso recorrente de antibióticos na infância aumenta o risco de doença inflamatória intestinal. Essa hipótese antecipa a direção do efeito, portanto é unilateral. Por que, então, usar uma hipótese alternativa bilateral ao planejar o tamanho de amostra? A resposta é que, na maioria das vezes, ambos os lados da hipótese alternativa (i. e., maior ou menor risco) são interessantes, havendo interesse em publicar os resultados, independentemente da

direção que foi observada no estudo. O rigor estatístico exige que o investigador escolha entre hipóteses uni e bilaterais antes de analisar os dados. Mudar de uma hipótese alternativa bilateral para uma hipótese unilateral para reduzir o valor *P* (veja mais adiante) não é um procedimento correto. Além disso, e esse é provavelmente o real motivo pelo qual hipóteses alternativas bilaterais são muito mais comuns, muitos revisores de solicitações para financiamento de pesquisa e de artigos submetidos para publicação esperam hipóteses bilaterais e não são receptivos a hipóteses unilaterais.

PRINCÍPIOS ESTATÍSTICOS BÁSICOS

Uma hipótese de pesquisa, como a de que quinze minutos ou mais de exercícios físicos por dia estão associados a uma média mais baixa de glicemia de jejum em mulheres de meia-idade com diabetes, pode ser verdadeira ou falsa no mundo real. Como o investigador não pode estudar todas as mulheres de meia-idade com diabetes, ele deve testar a hipótese em uma amostra da população-alvo. Conforme já foi mostrado na Figura 1.5, sempre haverá a necessidade de inferir sobre os fenômenos na população a partir de eventos observados na amostra. Infelizmente, devido ao acaso, às vezes o que ocorre em uma amostra não reflete o que teria ocorrido se toda a população tivesse sido estudada.

De uma certa forma, o problema do investigador é semelhante àquele enfrentado por um júri ao julgar um réu (Tabela 5.1). Em geral, é impossível determinar a verdade absoluta sobre se o réu cometeu o crime. Pelo contrário, o júri começa pressupondo a inocência, isto é, que o réu não cometeu o crime. O júri deve então decidir se há evidências suficientes para **rejeitar a inocência pressuposta** do réu. Esse padrão é conhecido no direito norte-americano como ***beyond reasonable doubt*** (**acima de uma dúvida razoável**). No entanto, o júri pode errar, ao condenar um réu inocente ou ao não condenar um réu culpado.

TABELA 5.1 Analogia entre as decisões do júri e os testes estatísticos

DECISÃO DO JÚRI	TESTE ESTATÍSTICO
Inocência: O réu não falsificou dinheiro	**Hipótese nula:** Não há associação entre o consumo de caroteno e a incidência de câncer de cólon na população
Culpa: O réu falsificou dinheiro	**Hipótese alternativa:** Há uma associação entre o consumo de caroteno e a

	incidência de câncer de cólon
Padrão para rejeitar a inocência: Acima de uma dúvida razoável	**Padrão para rejeitar uma hipótese nula:** Nível de significância estatística (α
Julgamento correto: Condenar um estelionatário	**Inferência correta:** Concluir que há associação entre o consumo de caroteno e câncer de cólon quando realmente houver essa associação na população
Julgamento correto: Absolver uma pessoa inocente	**Inferência correta:** Concluir que não há associação entre o consumo de caroteno e câncer de cólon quando realmente não houver essa associação
Julgamento incorreto: Condenar uma pessoa inocente	**Inferência incorreta (erro Tipo I):** Concluir que há associação entre o consumo de caroteno e câncer de cólon quando não houver associação
Julgamento incorreto: Absolver um estelionatário	**Inferência incorreta (erro Tipo II):** Concluir que não há associação entre o consumo de caroteno e câncer de cólon quando houver associação

Da mesma forma, o investigador começa pressupondo a hipótese nula de que não há associação entre as variáveis preditora e de desfecho na população. Com base nos dados coletados na amostra, ele usa testes estatísticos para determinar se há evidências suficientes para rejeitar a hipótese nula em benefício da hipótese alternativa de que há associação na população. O padrão para esses testes é conhecido como **nível de significância estatística**.

Erros Tipo I e Tipo II

Da mesma forma que ocorre com um júri, o investigador pode chegar a uma conclusão incorreta. Às vezes, uma amostra não é representativa da população tão somente pelo acaso. Quando isso ocorre, os resultados na amostra não refletem a realidade na população, levando a inferências errôneas. Um **erro tipo I** (falso-positivo) ocorre quando se rejeita uma hipótese nula que é verdadeira na população; um **erro tipo II** (falso-negativo) ocorre quando se deixa de rejeitar (aceita-se) uma hipótese nula que é falsa na população. Embora os erros tipo I e tipo II não possam ser totalmente evitados, é possível reduzir a probabilidade de sua ocorrência, aumentando-se o tamanho de amostra (quanto maior a amostra, menor a probabilidade de ela diferir substancialmente da realidade vivida pela população) ou ajustando-se o delineamento ou as aferições nos moldes que serão discutidos mais adiante.

Neste capítulo e no próximo, lidamos apenas com formas de reduzir os erros tipo I e tipo II que ocorrem devido à variação ao **acaso**, também

conhecido como erro aleatório. Os resultados falso-positivos e falso-negativos também podem ocorrer em função de viés, mas os erros decorrentes de vieses normalmente não são denominados erros tipo I e tipo II. Esses erros são mais complicados, por serem de difícil detecção e por não haver como quantificá-los com métodos estatísticos nem evitá-los aumentando-se o tamanho de amostra. (Veja Capítulos 1, 3, 4 e 7-12 sobre estratégias para a redução de erros causados por viés.)

Magnitude de efeito

A probabilidade de um estudo detectar uma associação entre a variável preditora e a de desfecho em uma amostra depende da magnitude real da associação na população-alvo. Se a associação for forte (p.ex., diferença de 20m g/dL na glicemia de jejum), ela será facilmente detectada na amostra. Por outro lado, se a associação for fraca (diferença de 2 mg/dL), será difícil detectá-la na amostra.

Infelizmente, quase nunca se conhece a magnitude da associação durante o planejamento da pesquisa, até mesmo porque um dos objetivos da pesquisa é estimá-la! Assim, o investigador precisa definir a magnitude da associação que ele deseja detectar na amostra. Esse valor é denominado **magnitude de efeito**. Definir adequadamente essa magnitude é o aspecto mais difícil do planejamento do tamanho de amostra (4). Deve-se tentar localizar dados de estudos anteriores em áreas afins para que se possa fazer um “chute informado” sobre a magnitude plausível do efeito esperado. Outra opção seria escolher uma magnitude mínima de efeito que poderia ser considerada clinicamente significativa (p. ex., uma redução de 10 mg/dL no nível de glicemia de jejum).

Sem dúvida, em termos de saúde pública, mesmo uma redução de 2 a 3 mg/dL nos níveis de glicemia de jejum pode ser importante, sobretudo se for facilmente alcançável. Portanto, a definição da magnitude de efeito é sempre arbitrária, e as considerações sobre factibilidade do estudo são fundamentais. Quando o número de sujeitos disponíveis ou acessíveis for limitado, o mais adequado é trabalhar de trás para frente (Capítulo 6), determinando a magnitude de efeito que o estudo poderá detectar, dado o número de sujeitos que poderão ser estudados.

Muitos estudos têm várias magnitudes de efeito, pois medem diversas variáveis preditoras e de desfecho. Ao delinear um estudo, deve-se

determinar o tamanho de amostra a partir da magnitude de efeito desejada para a hipótese mais importante. As magnitudes detectáveis dos demais efeitos podem, então, ser estimadas a partir desse tamanho de amostra. Caso existam várias hipóteses de mesma importância, o tamanho da amostra para o estudo deve basear-se na hipótese que exigir a maior amostra.

α, β e poder estatístico

Após o término de um estudo, o investigador usa testes estatísticos para tentar rejeitar a hipótese nula em benefício da hipótese alternativa, da mesma forma que um promotor tenta convencer o júri a rejeitar a inocência em benefício da culpa. Dependendo de se a hipótese nula é verdadeira ou falsa na população e pressupondo-se que o estudo esteja livre de vieses, quatro situações são possíveis (Tabela 5.2). Em duas delas, os achados na amostra e a verdade na população estão de acordo, e a inferência do investigador será correta. Nos outros dois casos, terá havido um erro tipo I ou tipo II, e a inferência será incorreta.

TABELA 5.2 Verdade na população vs. resultados na amostra do estudo: as quatro possibilidades

	VERDADE NA POPULAÇÃO	
RESULTADOS NA AMOSTRA DO ESTUDO	**ASSOCIAÇÃO ENTRE O PREDITOR E O DESFECHO**	**AUSÊNCIA DE ASSOCIAÇÃO ENTRE O PREDITOR E O DESFECHO**
Rejeitam a hipótese nula	Correta	Erro tipo I
Não rejeitam a hipótese nula	Erro tipo II	Correta

Antes de realizar o estudo, o investigador determina a probabilidade máxima tolerada para erros tipo I e tipo II. A probabilidade máxima de um erro tipo I (rejeitar a hipótese nula quando ela for verdadeira) é denominada α (**alfa**). Outro termo para a é **nível de significância estatística**.

Se, por exemplo, um estudo sobre os efeitos da atividade física nos níveis de glicemia de jejum for delineado com um a de 0,05, isso significa que se definiu como de 5% a probabilidade máxima de se rejeitar a hipótese nula quando ela for verdadeira (ou seja, inferir que há associação entre atividade física e níveis de glicemia de jejum na população, quando na verdade não há). Esse é o nível de incerteza que o investigador se

dispõe a aceitar quando for usar testes estatísticos para analisar os dados após o estudo ter sido concluído.

A probabilidade de um erro tipo II (não rejeitar a hipótese nula quando ela for falsa) é denominada β (**beta**). O valor $[1 - \beta]$ é denominado **poder estatístico** e representa a probabilidade de se rejeitar corretamente a hipótese nula na amostra quando o efeito real na população for igual (ou maior do que) a magnitude de efeito especificada.

Se β for 0,10, o investigador decidiu que está disposto a aceitar uma probabilidade de 10% de não encontrar uma associação de uma magnitude de efeito especificada, quando de fato essa associação existe. Isso representa um poder de 0,90, isto é, uma probabilidade de 90% de encontrar uma associação dessa magnitude ou maior. Por exemplo, imagine que o exercício físico realmente leve a uma redução de 20 mg/dL nos níveis de glicemia de jejum em mulheres diabéticas na população. Se o investigador replicasse o estudo várias vezes, com o mesmo poder estatístico de 90%, esperaríamos que, em 9 entre cada 10 estudos, ele rejeitaria corretamente a hipótese nula com um nível especificado de alfa de 0,05 e concluiria que o exercício físico está associado com a glicemia de jejum. Isso não significa que o investigador não poderia detectar um efeito menor, como uma redução de 15 mg/dL; significa apenas que teria menos de 90% de probabilidade de detectá-lo.

Em um cenário ideal, α e β seriam próximos de zero, minimizando a possibilidade de resultados falso-positivos e falso-negativos. Para reduzi-los, no entanto, é necessário aumentar o tamanho de amostra ou implementar outras estratégias discutidas no Capítulo 6. O planejamento do tamanho de amostra tem como meta escolher um número suficiente de sujeitos para manter α e β em níveis aceitavelmente baixos, sem tornar o estudo desnecessariamente caro ou trabalhoso.

Muitos estudos estabelecem a como 0,05 e β como 0,20 (poder de 0,80). Esses valores são arbitrários, sendo possível usar outros. Por exemplo, a faixa convencional para a varia de 0,01 α 0,10 e, para β, de 0,05 a 0,20. Em geral, o investigador deve usar um a baixo quando a questão de pesquisa torna importante evitar erros tipo I (falso-positivos), como ao testar a eficácia de um medicamento que apresenta riscos. Deve-se usar um β baixo (e uma magnitude de efeito pequena) quando for particularmente importante evitar um erro tipo II (falso-negativo), como

no caso em que se deseja fornecer evidências para garantir à opinião pública que é seguro viver próximo a um depósito de lixo tóxico.

Valor *P*

Vamos agora retornar à hipótese nula, cujo objetivo central ficará finalmente mais claro. A hipótese nula tem apenas uma única função: funcionar como um espantalho. É pressuposta como verdadeira até ser rejeitada como falsa por um teste estatístico. Na análise dos dados, um teste estatístico é usado para determinar o **valor *P***, que é a probabilidade de se encontrar – apenas pelo acaso – um efeito tão ou mais forte que o que seria encontrado no estudo se a hipótese nula fosse realmente verdadeira. O ponto central é perceber que, se a hipótese nula for verdadeira, e realmente não houver diferença na população, então a única forma que o estudo poderia ter encontrado uma diferença na amostra seria pelo acaso.

Se essa possibilidade for pequena, então a hipótese nula de que não há diferença pode ser rejeitada em benefício da hipótese alternativa de que há diferença. Por "pequena" queremos dizer que o valor *P* é menor que α, o nível predeterminado de significância estatística.

Entretanto, um resultado "**não significativo**" (valor *P* maior que α) não significa que não há associação na população; significa apenas que o resultado observado na amostra é pequeno comparado ao que poderia ser encontrado pelo mero acaso. Por exemplo, um investigador pode concluir que mulheres que participam de competições esportivas universitárias têm o dobro da probabilidade de serem submetidas à artroplastia de quadril quando mais velhas do que as que não participam dessas competições, mas, uma vez que o número artroplastias de quadril no estudo foi modesto, esse efeito aparentemente forte teve um valor *P* de apenas 0,08. Isso significa que, mesmo não havendo associação entre atividade esportiva e artroplastia de quadril na população, haveria, apenas pelo acaso, uma probabilidade de 8% de se encontrar uma associação de magnitude pelo menos semelhante à observada pelo investigador. Se o investigador tivesse configurado o nível de significância como um a bilateral de 0,05, teria concluído que a associação na amostra "não foi estatisticamente significativa".

Nesse caso, poderia ser tentado a mudar de opinião e alterar o valor *P*

para unilateral, relatando-o como “P = 0,04”. Uma opção melhor seria relatar que “Embora os resultados possam sugerir uma associação, não houve significância estatística (P = 0,08)”. Essa solução preserva a integridade do desenho original da hipótese bilateral e também reconhece que a significância estatística não é uma situação do tipo tudo ou nada.

Lados da hipótese alternativa

É importante lembrar que a hipótese alternativa tem, na verdade, dois lados, e um deles ou ambos podem ser testados na amostra usando-se **testes estatísticos uni ou bilaterais**[1]. Quando um teste estatístico bilateral é usado, o valor P inclui as probabilidades de cometer um erro tipo I em cada uma das direções, o que é aproximadamente duas vezes maior do que a probabilidade em apenas uma direção. Portanto, é fácil converter um valor P unilateral para um valor P bilateral, e vice-versa. Por exemplo, um valor P unilateral de 0,05 é geralmente igual a um valor P bilateral de 0,10. (Alguns testes estatísticos são assimétricos, e é por isso que empregamos a palavra “geralmente”.)

No caso incomum de um investigador estar apenas interessado em um dos lados da hipótese alternativa (p. ex., em um ensaio clínico de não inferioridade que busca determinar se um novo antibiótico não é menos eficaz do que um antibiótico atualmente em uso. Veja o Capítulo 11), o tamanho de amostra pode ser calculado de acordo com a hipótese unilateral. Uma hipótese unilateral, entretanto, nunca deve ser usada apenas para reduzir o tamanho de amostra.

Tipo de teste estatístico

As fórmulas usadas para calcular o tamanho de amostra são baseadas em pressupostos matemáticos que diferem para cada teste estatístico. Portanto, antes de se poder calcular o tamanho de amostra, o investigador deve decidir que método estatístico usar para a análise dos dados. Essa escolha depende principalmente dos tipos de variáveis preditoras e de desfecho no estudo. A Tabela 6.1 lista algumas estatísticas comuns usadas na análise de dados, e o Capítulo 6 fornece métodos simplificados para estimar o tamanho de amostra para estudos que usam essas estatísticas.

PONTOS ADICIONAIS

Variabilidade

Além da magnitude de um efeito, deve-se atentar também para a sua variabilidade. Os testes estatísticos dependem da capacidade de mostrar uma diferença entre os grupos comparados. Quanto maior a variabilidade (ou dispersão) na variável de desfecho entre os sujeitos, maior a probabilidade de os valores nos grupos se sobrecruzarem, e maior a dificuldade de se demonstrar uma diferença global entre eles. Uma vez que o erro na aferição contribui para a variabilidade global, medidas menos precisas exigem tamanhos de amostra maiores (5).

Considere um estudo sobre os efeitos de duas dietas (uma com baixos níveis de gordura e outra com baixos níveis de carboidratos) sobre a perda de peso em 20 pacientes obesos. Se todos os pacientes que fizerem uma dieta com baixos níveis de gordura perderem 3 kg e todos os que fizerem uma dieta com baixos níveis de carboidratos perderem pouco ou nenhum peso (uma magnitude de efeito de 3 kg), é provável que a dieta com baixos níveis de gordura seja melhor (Figura 5.1A). Por outro lado, se a média de perda de peso for de 3 kg no grupo com baixos níveis de gordura e de 0 kg no grupo com baixos níveis de carboidratos, mas houver grande sobreposição entre os dois grupos (o caso da Figura 5.1B), a variabilidade maior tornará mais difícil detectar uma diferença entre as duas dietas, sendo necessário um tamanho maior de amostra para demonstrar a diferença.

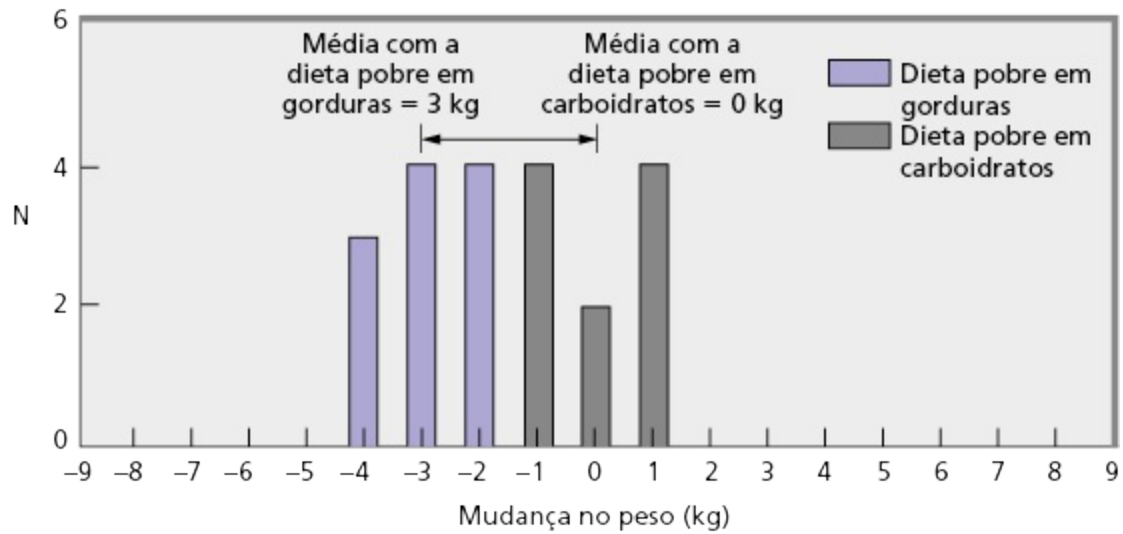
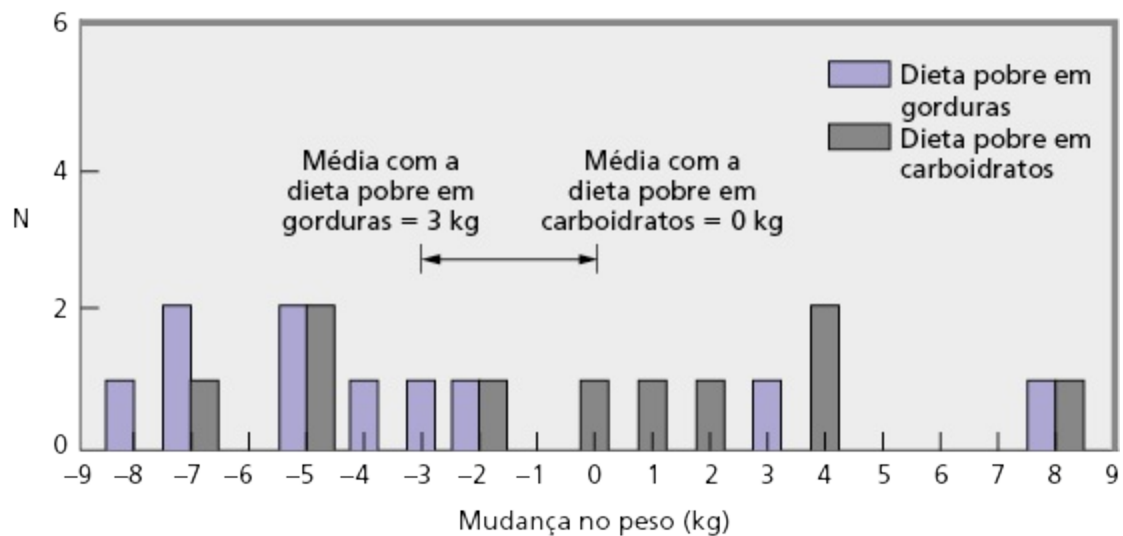



FIGURA 5.1 A: Perda de peso alcançada por duas dietas. Todos os sujeitos na dieta pobre em gordura perderam de 2 a 4 kg, ao passo que a mudança de peso no grupo da dieta pobre em carboidratos variou de –1 a +1 kg. Uma vez que não há sobreposição entre os dois grupos, é razoável inferir que a dieta pobre em gordura teve um desempenho melhor do que a dieta pobre em carboidratos (como seria confirmado por um teste *t* com um valor $P < 0{,}0001$). **B**: Perda de peso alcançada por duas dietas. Há grande sobreposição na mudança de peso nos dois grupos. Embora a magnitude do efeito seja a mesma (3 kg) como a em **A**, há pouca evidência de que uma dieta seja superior à outra (como seria confirmado por um teste *t*, que tem um valor *P* de 0,19).

Quando uma das variáveis usadas na estimativa do tamanho de amostra for contínua (p. ex., peso corporal na Figura 5.1), o investigador precisará estimar a sua variabilidade. (Veja item sobre o teste *t* no Capítulo 6 para

mais detalhes.) Nos demais casos, a variabilidade já está incluída nos outros parâmetros usados nas fórmulas e tabelas de tamanho de amostra e não precisa ser especificada.

Hipóteses múltiplas e *post hoc*

Quando mais de uma hipótese for testada em um estudo, especialmente se algumas das hipóteses forem formuladas após a análise dos dados (hipóteses *post hoc*), aumenta a probabilidade de que pelo menos uma delas irá alcançar significância estatística tão somente pelo acaso. Por exemplo, se 20 hipóteses independentes são testadas a um a de 0,05, há uma grande probabilidade (64%; [1 – $0{,}95^{20}$]) de que pelo menos uma hipótese será estatisticamente significativa tão somente pelo acaso. Certos estatísticos recomendam que se ajuste o nível de significância estatística no caso de mais de uma hipótese ser testada. Isso mantém a probabilidade global de se aceitar qualquer uma das hipóteses alternativas para o nível especificado de significância quando todos os achados se devem ao acaso. Por exemplo, estudos genômicos que procuram uma associação entre milhares de genótipos e doença precisam usar um a muito menor que 0,05 ou correm o risco de identificar muitas associações falso-positivas.

Um método, denominado **procedimento de Bonferroni**, em homenagem a esse matemático, é dividir o nível de significância estatística (p. ex., 0,05) pelo número de hipóteses testadas. Por exemplo, se houver quatro hipóteses, cada uma deve ser testada a um a de 0,0125 (0,05 ÷ 4). Isso requer um aumento considerável no tamanho de amostra se comparado ao necessário para se testar cada uma das hipóteses a um α de 0,05. Portanto, para qualquer hipótese, o método de Bonferroni reduz a probabilidade de um erro tipo I ao custo de aumentar a probabilidade de um erro tipo II ou de requerer um maior tamanho de amostra. Se os resultados de um estudo ainda forem estatisticamente significativos após o ajuste de Bonferroni, essa perda de poder não é um problema. No entanto, quando um resultado perde significância estatística após o ajuste de Bonferroni, isso poderia significar deixar de apoiar uma associação que realmente estava presente na população (um erro tipo II), o que é mais problemático.

Especialmente nesses casos, a questão sobre que nível de significância

usar depende mais da **probabilidade *a priori*** de cada hipótese que do número de hipóteses testadas e, por essa razão, nossa opinião geral é de que a abordagem insensata de Bonferroni para o teste de hipóteses múltiplas é, muitas vezes, exigente demais. Pode-se fazer uma analogia com a utilidade de usar testes diagnósticos (6, 7). Ao interpretar os resultados de um teste diagnóstico, um clínico considera a probabilidade de que o paciente testado tenha a doença em questão. Por exemplo, um resultado de um teste ligeiramente anormal em uma pessoa sadia (p. ex., um nível de fosfatase alcalina sérica 15% maior do que o limite superior do normal) é provavelmente um resultado falso-positivo, de pequena importância clínica. Da mesma forma, um valor *P* de 0,05 para uma hipótese improvável é, também, um resultado provavelmente falso-positivo.

No entanto, um nível de fosfatase alcalina 10 ou 20 vezes maior do que o limite superior do normal provavelmente não ocorreria apenas por acaso (embora possa ser um erro laboratorial). Assim, é pouco provável que um valor *P* muito pequeno (p. ex., < 0,001) também tenha ocorrido por acaso (embora possa ocorrer devido a um viés). É difícil descartar resultados de testes muito anormais como falso-positivos ou descartar valores *P* muito baixos considerando-os como devidos ao acaso, mesmo quando a probabilidade *a priori* da doença ou da hipótese for baixa.[2]

Além disso, o número de testes solicitados ou hipóteses testadas nem sempre é relevante. A interpretação de um nível elevado de ácido úrico sérico em um paciente com articulação dolorida e edemaciada não deveria depender do fato de o médico ter pedido um único teste (nível de ácido úrico) ou obtido o resultado como parte de uma bateria de 20 testes. Da mesma forma, ao interpretar o valor *P* para testar uma hipótese de pesquisa que faz sentido, não importa se o investigador também testou outras hipóteses menos prováveis. O que mais importa é a plausibilidade da hipótese de pesquisa sendo testada. Em outras palavras, que ela tenha uma grande probabilidade *a priori* de ser correta. (A probabilidade *a priori*, nessa abordagem “**bayesiana**”, é normalmente um juízo subjetivo fundamentado em evidências de outras fontes.) A maioria das hipóteses formuladas durante o delineamento de um estudo normalmente atende a essa exigência. Afinal, por que outro motivo o investigador iria despender tempo e esforço em planejar e implementar o estudo?

E quando surgem associações não antecipadas durante a coleta e análise dos resultados de um estudo? Esse processo é denominado **geração de hipótese** ou, em um tom menos favorável, "mineração de dados" (*data-mining*) ou "pescaria" (*fishing expedition*). As inúmeras comparações informais feitas durante a análise dos dados são uma forma de testar hipóteses múltiplas. Um problema semelhante ocorre quando as variáveis são redefinidas durante a análise dos dados ou quando os resultados de apenas alguns subgrupos da amostra são apresentados. Os valores *P* significativos para hipóteses que foram geradas a partir de dados, mas que não foram consideradas durante o delineamento do estudo, muito frequentemente se devem ao acaso. Devem ser vistos com ceticismo e considerados apenas como fonte de potenciais questões de pesquisa para estudos posteriores.

No entanto, às vezes o investigador deixa de especificar uma determinada hipótese antecipadamente, embora essa hipótese pareça plausível quando os dados são analisados. Isso pode ocorrer, por exemplo, quando outros pesquisadores descobrem um novo fator de risco durante a realização do estudo ou quando o investigador inicialmente não havia percebido que a hipótese era boa. Assim, o que importa não é se a hipótese foi formulada antes do início do estudo, mas se a probabilidade *a priori* da hipótese é sólida, com base em evidências de outras fontes sobre sua veracidade (6, 7).

Há vantagens claras em se formular mais de uma hipótese no planejamento do estudo. O uso de **múltiplas hipóteses não relacionadas** aumenta a eficiência do estudo, tornando possível responder a mais questões com um único esforço de pesquisa e descobrir mais associações verdadeiras na população. Formular várias hipóteses relacionadas também pode ser uma boa opção. Se os achados forem consistentes, fortalecem as conclusões da pesquisa. Vários estudos em pacientes com insuficiência cardíaca mostraram que o uso de inibidores da enzima conversora da angiotensina ajuda a reduzir hospitalização, a mortalidade cardiovascular e a mortalidade total. Se apenas uma dessas hipóteses tivesse sido testada, as inferências desses estudos seriam menos definitivas. No entanto, o custo de testar múltiplas hipóteses é, obviamente, mais elevado. Imagine que várias hipóteses pré-formuladas sejam testadas e delas apenas uma ou duas se mostrem estatisticamente significativas. Nesse caso, o

investigador deve decidir (e tentar convencer os revisores, editores e leitores) sobre os resultados significativos, os não significativos ou ambos os conjuntos de resultados são corretos.

Hipóteses principais e secundárias

Alguns estudos, especialmente grandes ensaios clínicos randomizados, especificam algumas hipóteses como "**secundárias**". Isso geralmente ocorre quando existe uma **hipótese principal** para a qual o estudo foi delineado, mas os investigadores também estão interessados em outras questões de pesquisa de menor importância. Por exemplo, o desfecho principal em um ensaio clínico sobre suplementação de zinco pode ser hospitalizações ou visitas à emergência para tratar infecções das vias aéreas superiores, e um desfecho secundário poderia ser o autorrelato dos dias perdidos no trabalho ou na escola. Se o estudo está sendo conduzido para obter a aprovação de um medicamento, então o desfecho principal é o que realmente importa para o órgão regulador. Uma hipótese secundária postulada previamente aumenta a credibilidade dos resultados quando essa hipótese for testada.

Uma boa regra, especialmente para ensaios clínicos, é estabelecer antecipadamente tantas hipóteses quanto façam sentido, mas especificar apenas uma como **hipótese principal**, que poderá ser testada estatisticamente, sem preocupação quanto à necessidade de ajustar para hipóteses múltiplas. E, o que é mais importante, ter uma hipótese principal ajuda também no direcionamento do enfoque do estudo para seu objetivo principal e fornece um suporte claro para o cálculo principal do tamanho de amostra necessário.

Muitos estatísticos e epidemiologistas estão se afastando do teste de hipóteses, com sua ênfase em valores *P*, e passando a usar mais os intervalos de confiança para relatar a precisão dos resultados do estudo (8-10). De fato, alguns autores acreditam que todo o processo de basear o planejamento do tamanho da amostra em hipóteses é enganador, em parte porque depende de valores que são desconhecidos (magnitude do efeito) ou arbitrários (α e β) (11). Entretanto, a abordagem que descrevemos é prática e continua sendo a mais utilizada no planejamento da pesquisa clínica.

RESUMO

1. O **planejamento do tamanho de amostra** é uma parte importante do delineamento de estudos analíticos e descritivos. O tamanho de amostra deve ser estimado em um estágio inicial no processo de delineamento da pesquisa, de forma que mudanças adicionais que se façam necessárias ainda possam ser implementadas.
2. Os estudos analíticos e os experimentais necessitam de uma **hipótese** que especifique, para fins de **testes de significância** posteriores, a associação antecipada entre as variáveis preditoras e de desfecho principais. Os estudos puramente descritivos, que não apresentam estratégia de comparação, não exigem a formulação de uma hipótese.
3. Boas hipóteses são **específicas** em termos de como a população será amostrada e como as variáveis serão medidas, **simples** (com apenas uma variável preditora e uma de desfecho) e **formuladas em um estágio inicial**.
4. A **hipótese nula**, que propõe que a variável preditora não está associada com o desfecho, é a base para os testes de significância estatística. A **hipótese alternativa** propõe que há associação. Os testes estatísticos tentam rejeitar a hipótese nula de que não há associação em benefício da hipótese alternativa de que há associação.
5. A hipótese alternativa é **unilateral** (apenas uma direção da associação será testada) ou **bilateral** (ambas as direções serão testadas). As hipóteses unilaterais somente devem ser usadas em casos raros quando apenas uma direção da associação for clínica ou biologicamente relevante.
6. Para estudos analíticos e experimentais, o tamanho de amostra é uma estimativa do número de sujeitos necessários para detectar uma associação de uma determinada **magnitude de efeito** e **variabilidade**, com uma probabilidade especificada de incorrer em **erros tipo I** (falso-positivo) e **tipo II** (falso-negativo). A probabilidade máxima de um erro tipo I é denominada a; a de um erro tipo II, β O valor (1 – β) é o **poder estatístico**, a probabilidade de se observar uma associação de uma determinada magnitude ou maior em uma amostra se essa associação realmente existir na população.
7. Muitas vezes é desejável estabelecer mais de uma hipótese em um

estágio inicial, mas o investigador deve especificar uma única **hipótese principal** como foco de delineamento e para a estimativa de tamanho de amostra. A interpretação dos achados a partir de testes de **hipóteses múltiplas** na amostra, incluindo os achados não antecipados que emergirem dos dados, deve basear-se em juízo sobre a **probabilidade *a priori*** de que eles representem um evento real na população.

REFERÊNCIAS

1. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 1994;330:1029–1035.
2. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 1991;324:781–788.
3. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* 1992;327:227–233.
4. Van Walraven C, Mahon JL, Moher D, et al. Surveying physicians to determine the minimal important difference: implications for sample-size calculation. *J Clin Epidemiol* 1999;52:717–723.
5. McKeown-Eyssen GE, Tibshirani R. Implications of measurement error in exposure for the sample sizes of casecontrol studies. *Am J Epidemiol* 1994;139:415–421.
6. Browner WS, Newman TB. Are all significant Pvalues created equal? The analogy between diagnostic tests and clinical research. *JAMA* 1987;257:2459–2463.
7. Newman TB, Kohn, MA. *Evidence-based diagnosis*. New York: Cambridge University Press, 2009. Chapter 11.
8. Daly LE. Confidence limits made easy: interval estimation using a substitution method. *Am J Epidemiol* 1998;147:783–790.
9. Goodman SN. Toward evidence-based medical statistics. 1: The Pvalue fallacy. *Ann Intern Med* 1999;130:995–1004.
10. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 1999;130:1005–1013.
11. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. *BMC Med*. 2010;8:17.

---

[1]Esses testes são às vezes referidos como uni e bicaudais, devido às caudas (extremidades) das distribuições estatísticas.

[2]Novamente, a exceção é alguns estudos genéticos, nos quais milhões ou mesmo bilhões de associações podem ser examinadas.

CAPÍTULO 6

Estimando o tamanho de amostra e o poder estatístico: aplicações e exemplos

Warren S. Browner, Thomas B. Newman e Stephen B. Hulley

O Capítulo 5 introduziu os princípios básicos que norteiam os cálculos de tamanho de amostra. Este capítulo apresenta várias receitas para o uso desses princípios para estimar o tamanho de amostra necessário para um projeto de pesquisa. A primeira seção expõe como estimar o **tamanho de amostra em estudos analíticos ou experimentais**, incluindo tópicos especiais com aplicabilidade a esses estudos, como análise multivariada. A segunda seção trata de **estudos primariamente descritivos**. As demais seções abordam estudos com **tamanho de amostra fixo**, estratégias para **maximizar o poder estatístico** e procedimentos para estimar o tamanho de amostra quando parece **não haver informações suficientes** para fazê-lo. O capítulo encerra com a apresentação de **erros comuns** a serem evitados.

Nos apêndices, são apresentadas tabelas e fórmulas para vários métodos básicos de estimativa de tamanho de amostra. Além disso, há uma calculadora na página deste livro na internet (em inglês) (*www.epibiostat.ucsf.edu/dcr/*), e existem muitas outras páginas na internet que possibilitam cálculos de tamanho de amostra interativos e instantâneos; faça uma busca por "*sample size calculator*" (calculadora de tamanho de amostra). A maioria dos pacotes estatísticos também pode estimar o tamanho de amostra para delineamentos de pesquisa comuns.

TÉCNICAS DE TAMANHO DE AMOSTRA PARA ESTUDOS ANALÍTICOS E EXPERIMENTAIS

Há inúmeras variações na receita de como estimar o tamanho de amostra em um estudo analítico ou experimental, mas todas apresentam passos em comum:

1. Definir a **hipótese nula** e uma **hipótese alternativa uni** ou **bilateral**.
2. Selecionar o **teste estatístico** apropriado na Tabela 6.1 com base no tipo de variável preditora e de desfecho dessas hipóteses.
3. Definir uma **magnitude de efeito** adequada (e uma **variabilidade**, se necessário).
4. Estabelecer α e β. (Especificar um α bilateral, a não ser que a hipótese alternativa seja evidentemente unilateral.)
5. Usar a tabela ou fórmula apropriada do apêndice, uma calculadora online ou um pacote estatístico para estimar o tamanho de amostra.

TABELA 6.1 Testes estatísticos simples para usar na estimativa do tamanho de amostra*

	VARIÁVEL DE DESFECHO	
VARIÁVEL PREDITORA	**DICOTÔMICA**	**CONTÍNUA**
Dicotômica	Teste do qui-quadrado†	Teste *t*
Contínua	Teste *t*	Coeficiente de correlação

* Veja mais adiante neste capítulo, na seção "Outras Considerações e tópicos especiais" o que fazer em relação a variáveis ordinais, ou se estiver planejando analisar os dados com outro tipo de teste estatístico

† O teste do qui-quadrado é sempre bilateral; um equivalente unilateral é a estatística Z.

Mesmo se o valor exato para um ou mais dos ingredientes for incerto, é importante estimar o tamanho de amostra cedo na fase de delineamento. Esperar até o último minuto para fazê-lo pode ser desastroso. Muitas vezes, é necessário começar tudo de novo, com novos ingredientes, o que pode significar redelinear todo o estudo. É justamente por isso que esse tema é tratado tão cedo neste livro.

Nem todos os estudos analíticos se enquadram perfeitamente em uma das três categorias principais de cálculo de tamanho de amostra descritas nas seções a seguir – teste do qui-quadrado quando a variável preditora e a de desfecho são ambas dicotômicas; teste *t* quando uma é dicotômica e a outra, contínua; e coeficiente de correlação quando ambas são contínuas. Algumas das exceções mais comuns são discutidas mais adiante neste

capítulo, na seção “Outras considerações e tópicos especiais”.

Teste *t*

O **teste *t*** (às vezes chamado de teste *t* de Student, pseudônimo de seu criador) é comumente usado para determinar se o valor médio de uma variável contínua em um grupo difere significativamente do de outro grupo. Por exemplo, o teste *t* seria apropriado para comparar as médias dos escores de depressão em pacientes tratados com dois antidepressivos diferentes, ou as médias do índice de massa corporal em pessoas com e sem diabetes. O teste *t* pressupõe que a distribuição da variável em cada um dos dois grupos se aproxima de uma curva normal (em forma de sino). No entanto, por ser um teste extremamente robusto, pode ser usado para praticamente qualquer distribuição, a não ser que o número de sujeitos seja pequeno (menos de 30 ou 40) ou que estejam presentes valores extremos (*outliers*, ou valores atípicos).

Embora o teste *t* seja geralmente usado para comparar desfechos contínuos, também pode ser usado para estimar o tamanho de amostra para um desfecho dicotômico (p. ex., em um estudo de caso-controle) se o estudo tiver uma variável preditora contínua. Nesse caso, o teste *t* compara o valor médio da variável preditora entre os casos e os controles.

Para estimar o tamanho de amostra para um estudo no qual os valores médios de uma variável de desfecho contínua serão comparados usando um teste *t* (veja Exemplo 6.1), o investigador deve:

1. Formular a hipótese nula e decidir se a hipótese alternativa é uni ou bilateral.
2. Estimar a magnitude de efeito (*E*) a partir da diferença no valor médio da variável contínua entre os grupos de estudo.
3. Estimar a variabilidade como o desvio-padrão (*DP*) da variável.
4. Calcular a magnitude padronizada de efeito (*E*/*DP*), definida como o quociente entre a magnitude de efeito e o desvio-padrão da variável de desfecho.
5. Estabelecer α e β.

EXEMPLO 6.1 Tamanho de amostra quando se usa o teste *t*

Problema: Para a questão de pesquisa “há diferença entre a eficácia

do salbutamol e a do brometo de ipratrópio no tratamento da asma?", o investigador planeja um ensaio clínico randomizado sobre o efeito desses medicamentos no VEF_1 (volume expiratório forçado no primeiro segundo) após duas semanas de tratamento. Um estudo anterior apontou que o VEF_1 médio em indivíduos com asma tratada foi de 2,0 L, com um desvio-padrão de 1,0 L. O investigador gostaria de poder detectar uma diferença de 10% ou mais no VEF_1 médio entre os dois grupos de tratamento. Quantos pacientes seriam necessários em cada grupo (salbutamol e ipratrópio) para α (bilateral) = 0,05 e poder = 0,80?

Os ingredientes para o cálculo do tamanho de amostra são apresentados a seguir:

1. **Hipótese nula:** o VEF_1 médio após duas semanas de tratamento é o mesmo em pacientes asmáticos tratados com salbutamol e nos tratados com ipratrópio.
   Hipótese alternativa (bilateral): o VEF_1 médio após duas semanas de tratamento em pacientes asmáticos tratados com salbutamol é diferente daquele de pacientes tratados com ipratrópio.
2. Magnitude de efeito = 0,2 L (10% × 2,0 L).
3. Desvio-padrão do VEF_1 = 1,0 L.
4. Magnitude padronizada de efeito = magnitude de efeito ÷ desvio-padrão = 0,2 L ÷ 1,0 L = 0,2.
5. α (bilateral) = 0,05; β = 1 – 0,80 = 0,20. (Lembre-se de que β = 1 – poder.)

Partindo de uma magnitude padronizada de efeito (coluna da esquerda) de 0,20 na Tabela 6A e movendo para a direita até um α (bilateral) = 0,05 e um β = 0,20, são necessários 394 pacientes por grupo. Esse seria o número de pacientes necessários ao final do estudo, no entanto, como muitos abandonam o estudo antes de ele ser completado, é preciso arrolar um número maior que 394. Esse tamanho de amostra poderá não ser factível, tornando necessário revisar o delineamento da pesquisa ou, talvez, aceitar que somente será possível detectar uma magnitude maior de efeito. Veja a seção sobre teste *t* para amostras pareadas ("Exemplo 6.8") para uma

potencial solução para esse problema.

A **magnitude de efeito** e a **variabilidade** costumam ser estimadas a partir de dados de estudos anteriores publicados na literatura e de consultas a especialistas. Eventualmente, um pequeno estudo-piloto poderá ser necessário para estimar o desvio-padrão da variável (veja também a seção “Como estimar o tamanho de amostra quando as informações são insuficientes”, mais adiante, neste capítulo). Quando uma variável de desfecho for a mudança em uma medida contínua (p. ex., mudança de peso durante um estudo), deve-se usar o desvio-padrão da mudança nessa variável (não o desvio-padrão da própria variável) nas estimativas de tamanho de amostra. O desvio-padrão da mudança em uma variável é geralmente menor do que o desvio-padrão da variável; portanto, o tamanho de amostra será também menor.

Às vezes um investigador não consegue obter informações adequadas sobre o desvio-padrão de uma variável. Nesse caso, recomenda-se usar a **magnitude padronizada de efeito**, que é um valor sem unidade que permite estimar o tamanho de amostra; também simplifica comparações entre magnitudes de efeito de variáveis diferentes. A magnitude padronizada de efeito é simplesmente o quociente entre a magnitude de efeito e o desvio-padrão da variável. Por exemplo, uma diferença de 10 mg/dL no colesterol sérico, que tem um desvio-padrão na população de aproximadamente 40 mg/dL, tem uma magnitude padronizada de efeito de 0,25. Quanto maior a magnitude padronizada de efeito, menor o tamanho de amostra necessário. As magnitudes padronizadas de efeito para a maioria dos estudos serão $> 0,1$; magnitudes menores são difíceis de detectar (exigem grandes tamanhos de amostra) e em geral não são clinicamente importantes.

O Apêndice 6A fornece as exigências de tamanho de amostra para várias combinações de α e β e para várias magnitudes padronizadas de efeito. Para usar a Tabela 6A, identifique a magnitude padronizada de efeito na coluna mais à esquerda, movendo-se para a direita até os valores definidos para α e β, o que leva ao tamanho de amostra exigido por grupo. (Os números da Tabela 6A pressupõem que os dois grupos que estão sendo comparados têm o mesmo tamanho; se esse pressuposto não for verdadeiro, utilize a fórmula abaixo da tabela, um pacote estatístico ou

uma calculadora interativa na internet.)

Um bom **macete** para estimar o tamanho aproximado de amostra usando o teste *t*, quando serão estudados mais de 30 sujeitos e o poder for estabelecido como 0,80 (β = 0,2) e o α (bilateral) como 0,05 (1), é empregar a seguinte fórmula simplificada:

Tamanho da amostra (por grupo de igual tamanho) = 16 ÷ (magnitude padronizada de efeito)^2

Para o Exemplo 6.1, a estimativa de tamanho de amostra obtida por essa fórmula seria 16 ÷ 0,22 = 400 por grupo.

Teste do qui-quadrado

O **teste do qui-quadrado (X^2)** é usado para comparar a proporção de sujeitos em cada um de dois grupos que apresentam um desfecho dicotômico. Por exemplo, a proporção de homens que desenvolvem doença arterial coronariana (DAC) quando tratados com ácido fólico pode ser comparada com a proporção de homens que desenvolvem DAC quando recebem placebo. O teste do qui-quadrado é sempre bilateral; um teste equivalente para hipóteses unilaterais é o **teste *Z* unilateral.**

Em um ensaio clínico ou estudo de coorte, a magnitude de efeito é especificada pela diferença entre P_1 (proporção esperada de sujeitos que apresentam o desfecho em um grupo, isto é, o risco de desenvolver o desfecho) e P_2 (proporção esperada no outro grupo). Por exemplo, em um estudo de coorte que compara o risco de desenvolver doença renal terminal em homens e mulheres com hipertensão, P_1 seria a proporção de homens que desenvolvem doença renal terminal e P_2 seria a proporção de mulheres que desenvolvem esse mesmo desfecho. A variabilidade é uma função de P_1 e P_2, e, portanto, não precisa ser especificada.

Por outro lado, quando se calcula o tamanho de amostra para um estudo de caso-controle, P_1 *e* P_2 têm outras definições. Eles se referem às proporções esperadas de casos e controles com um determinado valor de uma variável preditora dicotômica (p. ex., a proporção de casos de doença renal terminal que eram homens). Assim, em um estudo de caso-controle, P_1 representa a proporção esperada para um determinado valor de uma

variável dicotômica entre os casos (isto é, a frequência desse valor entre os casos), e P_2 representa a proporção esperada para esse valor entre os controles.

Para estimar o tamanho de amostra para um estudo que será analisado com o teste do qui-quadrado ou com o teste Z para comparar duas proporções (Exemplo 6.2), o investigador deve:

1. Formular a hipótese nula e decidir se a hipótese alternativa é uni ou bilateral.
2. Estimar a magnitude de efeito e a variabilidade em termos de P_1 (proporção com o desfecho em um grupo) e P_2 (proporção com o desfecho no outro grupo).
3. Estabelecer α e β.

EXEMPLO 6.2 Cálculo do tamanho de amostra para o teste do qui-quadrado

Problema: A questão de pesquisa é: "Pessoas que praticam Tai Chi têm um menor risco de desenvolver dor lombar do que as que praticam corrida?" Uma revisão da literatura sugere que o risco de desenvolver dor lombar em dois anos é de em torno de 0,30 em pessoas que praticam corrida. O investigador espera ser capaz de mostrar que o Tai Chi reduz o risco em pelo menos 0,10. Para um α (bilateral) = 0,05 e poder = 0,80, quantas pessoas deverão ser estudadas para se determinar se a incidência de dor lombar em 2 anos é de 0,20 (ou menos) nas pessoas que praticam Tai Chi?

Solução: Os ingredientes para o cálculo do tamanho de amostra são os seguintes:

1. **Hipótese nula:** a incidência de dor lombar é a mesma em pessoas que correm e naquelas que praticam Tai Chi.
   Hipótese alternativa (bilateral): a incidência de dor lombar é diferente em pessoas que correm e naquelas que praticam Tai Chi.
2. P_2 (incidência em pessoas que correm) = 0,30; P_1 (incidência em pessoas que praticam Tai Chi) = 0,20. O menor entre esses valores é 0,20, e a diferença entre eles ($P_1 - P_2$) é 0,10.
3. α (bilateral) = 0,05; β = 1 – 0,80 = 0,20.

Partindo de 0,20 na coluna da esquerda na Tabela 6B.1 e de uma diferença esperada de 0,10, o valor central para α (bilateral) = 0,05 e β = 0,20 é de 313 pessoas que praticam corrida e 313 pessoas que praticam Tai Chi (esse é o tamanho de amostra exigido).

O Apêndice 6B apresenta os tamanhos de amostra necessários para várias combinações de α e β e para uma faixa de valores de P_1 e P_2. Para estimar o tamanho de amostra, localize na coluna da esquerda da Tabela 6B.1 ou 6B.2 o menor entre P_1 e P_2 (se necessário, arredondando para 0,05). Então, procure a diferença entre P_1 e P_2. A tabela fornece tamanhos de amostra necessários por grupo para os valores de α e β especificados.

Muitas vezes, especifica-se a magnitude de efeito em termos do **risco relativo** (razão de risco) para o desfecho nos dois grupos de estudo. Por exemplo, um investigador poderia estudar se mulheres usuárias de anticoncepcionais orais têm risco pelo menos duas vezes maior de sofrerem infarto do miocárdio do que as não usuárias. Em um estudo de coorte (ou um ensaio clínico), é muito simples converter risco relativo nas duas proporções (P_1 e P_2) e vice-versa, visto que o risco relativo é apenas P_1 dividido por P_2 (ou vice-versa).

Para um estudo de caso-controle, no entanto, a situação é um pouco mais complexa, pois o risco relativo deve ser aproximado pela **razão de chances (RC)** (*odds ratio*):

$$RC = \frac{[P_1 \times (1 - P_2)]}{[P_2 \times (1 - P_1)]}$$

O investigador deve especificar a razão de chances (RC) e o P_2 (proporção de controles expostos à variável preditora). Então o P_1 (proporção de casos expostos à variável preditora) pode ser calculado da seguinte forma:

$$P_1 = \frac{RC \times P_2}{(1 - P_2) + (RC \times P_2)}$$

Por exemplo, se o investigador espera que 10% dos controles usem anticoncepcionais orais (P_2 = 0,1) e deseja detectar uma razão de chances

de 3 associada a essa exposição, então:

$$P_1 = \frac{(3 \times 0,1)}{(1 - 0,1) + (3 \times 0,1)} = \frac{0,3}{1,2} = 0,25$$

Coeficiente de correlação

Embora o **coeficiente de correlação (*r*)** não seja comumente usado no cálculo do tamanho de amostra, ele pode ser usado quando as variáveis preditora e de desfecho são contínuas. O coeficiente de correlação é uma medida da força da associação linear entre as duas variáveis. Ele varia de –1 a 11. Valores negativos indicam que, à medida que uma variável aumenta, a outra diminui (como o nível de chumbo sérico e o QI de crianças). Quanto mais próximo o valor absoluto de *r* for de 1, mais forte será a associação; quanto mais próximo de 0, mais fraca a associação. A altura e o peso de adultos, por exemplo, estão altamente correlacionados em algumas populações, com $r \sim 0,9$. Tais valores altos, no entanto, são pouco comuns; muitas associações biológicas têm coeficientes de correlação muito menores.

Embora o coeficiente de correlação seja frequentemente usado em alguns campos da pesquisa clínica, tais como medicina comportamental, usá-lo para estimar o tamanho de amostra tem uma desvantagem: os coeficientes de correlação têm pouco significado intuitivo. Quando elevado ao quadrado (r^2), um coeficiente de correlação representa a proporção da dispersão (variância) em uma variável de desfecho que pode ser explicada por sua associação linear com uma variável preditora (e vice-versa). É por esta razão que valores de *r* pequenos, tais como $r \leq 0,3$, podem ser estatisticamente significativos para amostras suficientemente grandes, mas não ser clínica ou cientificamente importantes, pois explicam no máximo 9% da variância.

Uma forma alternativa (e muitas vezes a preferida) de estimar o tamanho de amostra para um estudo no qual as variáveis preditora e de desfecho são contínuas é dicotomizar uma das duas variáveis (digamos, na sua mediana) e usar, então, os cálculos do teste *t*. Essa estratégia tem a vantagem de expressar a magnitude de efeito como uma “diferença” entre dois grupos (a interpretação dos coeficientes de correlação, que não expressam a magnitude de efeito, é mais vaga). Para estimar o tamanho de

amostra em um estudo a ser analisado por coeficiente de correlação, o investigador deve:

1. Formular a hipótese nula e decidir se a hipótese alternativa é uni ou bilateral.
2. Estimar a magnitude de efeito como o valor absoluto do menor coeficiente de correlação (*r*) que o investigador gostaria de poder detectar. (A variabilidade é função de *r* e, como tal, já está incluída na tabela e na fórmula do apêndice.)
3. Estabelecer α e β.

No Apêndice 6C, localize na coluna da esquerda da Tabela 6C a magnitude do efeito (*r*) desejada. Então, localize horizontalmente, para os α e β. estabelecidos, o tamanho total de amostra exigido. A Tabela 6C apresenta o tamanho de amostra apropriado quando se quer rejeitar a hipótese nula de que não há associação entre as variáveis preditora e de desfecho (p. ex., $r = 0$). Se o objetivo for determinar se o coeficiente de correlação no estudo difere de um valor que não o zero (p. ex., $r = 0,4$), deve-se ler o texto abaixo da Tabela 6C para a metodologia apropriada.

EXEMPLO 6.3 Cálculo do tamanho de amostra para um coeficiente de correlação em um estudo transversal

Problema: Considere a questão de pesquisa “Os níveis de cotinina urinária (uma medida da intensidade do nível de fumo atual) estão correlacionados com a densidade óssea em fumantes?”. Um estudo anterior encontrou correlação modesta ($r = -0,3$) entre fumo relatado (em cigarros por dia) e densidade óssea (em g/cm^3); é possível antecipar que os níveis de cotinina urinária terão, no mínimo, uma correlação semelhante. Quantos fumantes deverão ser incluídos para α (bilateral) = 0,05 e β = 0,10?

Solução: Os ingredientes para o cálculo do tamanho de amostra são apresentados a seguir:

1. **Hipótese nula:** não há correlação entre nível de cotinina urinária e densidade óssea em fumantes.
   Hipótese alternativa: há correlação entre nível de cotinina urinária e densidade óssea em fumantes.

2. Magnitude de efeito (r) = | – 0,3| = 0,3.
3. α (bilateral) = 0,05; β = 0,10.

Na Tabela 6C, partindo de um r = 0,30, na coluna da esquerda, move-se horizontalmente até α (bilateral) = 0,05 e β = 0,10, onde se encontra o número de fumantes necessário: 113.

OUTRAS CONSIDERAÇÕES E TÓPICOS ESPECIAIS

Abandonos

Todas as unidades amostrais devem estar disponíveis para análise; assim, os sujeitos arrolados para o estudo, mas cujos desfechos não podem ser determinados (como em casos de **abandono**), não contam no tamanho de amostra. Se o investigador antecipar que alguns sujeitos não estarão disponíveis para o seguimento (o que geralmente é o caso), ele deverá estimar a proporção que será perdida e aumentar o tamanho da amostra arrolada de acordo com essa estimativa. Se, por exemplo, o investigador estimar que 20% de sua amostra será perdida no seguimento, o tamanho de amostra deverá ser aumentado por um fator de (1 ÷ [1 – 0,20]), ou 1,25.

Variáveis categóricas

Embora haja razões matemáticas pelas quais não seria adequado estimar um tamanho de amostra para **variáveis ordinais** usando um teste estatístico, na prática as variáveis ordinais podem muitas vezes ser tratadas como variáveis contínuas, especialmente se o número de categorias for relativamente grande (seis ou mais) e se fizer sentido calcular a média dos valores da variável.

Em outras situações, a melhor estratégia é mudar levemente a hipótese de pesquisa dicotomizando a variável categórica. Como exemplo, suponha que um pesquisador esteja estudando se falar inglês como segunda língua está associado ao número de consultas anuais de um paciente com diabetes com um especialista em pé diabético. O número de consultas é distribuído de forma desigual: muitos não farão consulta, outros farão uma única consulta, e apenas alguns farão duas ou mais consultas. Nesse caso, seria possível estimar o tamanho de amostra como

se o desfecho fosse dicotômico (nenhuma consulta *versus* uma ou mais consultas).

Análise de sobrevivência

Quando um investigador quer comparar a sobrevida ou outros dados do tipo tempo até o evento, como qual de dois tratamentos é mais eficaz em prolongar a vida em mulheres com câncer de mama avançado, a técnica apropriada para analisar os dados é a análise de sobrevivência (2,3). Embora a variável de desfecho, por exemplo meses de sobrevida, pareça ser contínua, o teste *t* não é adequado nessa situação. Isso porque o que está sendo avaliado, na verdade, não é o tempo propriamente dito (uma variável contínua), mas a proporção de sujeitos que permanecem vivos em cada momento (uma variável dicotômica). De forma semelhante, um investigador poderia comparar a taxa de desenvolvimento de um desfecho (por 100 pessoas-ano de seguimento) em dois grupos. Uma aproximação razoável pode ser feita simplesmente estimando as proporções esperadas de sujeitos com o desfecho nos dois grupos e estimando-se o tamanho de amostra com o teste do qui-quadrado. No entanto, caso se espere que o desfecho ocorra na maioria dos pacientes, como morte em um estudo sobre pacientes com câncer de mama avançado, uma estratégia melhor (uma vez que minimiza o tamanho total da amostra) seria fazer a estimativa com base nas proporções esperadas de sujeitos que terão desenvolvido o desfecho em cada grupo no momento em que aproximadamente metade de todos os desfechos tiverem ocorrido. Por exemplo, em um estudo que compara a sobrevida livre de doença em pacientes com câncer de mama tratadas com tratamento padrão *versus* experimental, no qual se espera que aproximadamente 60% das pacientes no tratamento-padrão terão morrido em dois anos, comparado com 40% das que receberam tratamento experimental, o tamanho da amostra pode ser estimado usando a “sobrevida em dois anos” como desfecho dicotômico.

Amostragem por conglomerados

Certos delineamentos de pesquisa apresentam **amostragem por conglomerados**, onde os sujeitos são amostrados por grupos (Capítulo 11). Considere, por exemplo, um estudo sobre se uma intervenção de

educação continuada para médicos melhora a taxa de cessação do tabagismo entre seus pacientes. Suponha que 20 clínicas sejam aleatoriamente alocadas ao grupo que recebe a intervenção e que 20 clínicas sejam alocadas a um grupo-controle. Os investigadores planejam revisar, um ano depois, os prontuários de uma amostra aleatória de 50 pacientes fumantes na linha de base de cada uma das clínicas e determinar quantos deixaram de fumar. Nesse caso, o tamanho de amostra seria 40 (número de clínicas nos dois grupos) ou 2.000 (número total de pacientes)? A resposta correta situa-se entre esses dois extremos e depende da semelhança entre os pacientes de cada clínica (quanto à probabilidade de parar de fumar) em comparação com a semelhança entre todos os pacientes. Para fazer essa estimativa, podem ser necessários dados de estudos-piloto, a não ser que outro investigador já tenha feito um estudo semelhante. Há várias técnicas para estimar o tamanho de amostra exigido para um estudo que utiliza amostragem por conglomerados (4-7), mas elas são de difícil uso e geralmente requerem o auxílio de um estatístico.

Pareamento (emparelhamento)

Vários motivos diferentes levam o investigador a optar por um delineamento pareado (Capítulo 9). As técnicas apresentadas neste capítulo, que ignoram qualquer pareamento, fornecem estimativas razoáveis do tamanho de amostra exigido, a não ser que a exposição (em um estudo de caso-controle pareado) ou o desfecho (em um estudo de coorte pareado) estejam fortemente correlacionados com a variável que está sendo usada para o pareamento. Estimativas mais precisas, que requerem que o investigador especifique a correlação entre as exposições ou desfechos nos pares que estão emparelhados, podem ser feitas usando-se abordagens-padrão (8), softwares estatísticos ou programas interativos disponíveis na internet.

Ajuste multivariado e outras análises estatísticas especiais

Muitas vezes, ao delinear-se um estudo observacional, conclui-se que uma ou mais variáveis poderão confundir a associação entre preditor e desfecho (Capítulo 9). Nesses casos, é possível incluir técnicas estatísticas no planejamento da análise dos resultados para ajustar para esses

confundidores. Quando esse ajuste for incluído no teste da hipótese principal, a estimativa do tamanho de amostra deve levar em conta esse fato.

Abordagens analíticas que ajustam para as variáveis confundidoras muitas vezes requerem maior tamanho de amostra (9,10). A magnitude do aumento depende de diversos fatores, incluindo a prevalência do confundidor, a força da associação entre preditor e confundidor e a força da associação entre confundidor e desfecho. Esses efeitos são complexos e não há regra que englobe todas as situações.

Foram desenvolvidos métodos multivariados, como regressão linear e regressão logística, que permitem ajustar para variáveis confundidoras. Outra técnica estatística bastante usada é a análise de **azares proporcionais de Cox**, que pode ajustar para confundidores e para diferenças de duração do seguimento. Se uma dessas técnicas for usada para analisar os dados, abordagens correspondentes podem ser usadas para estimar o tamanho de amostra exigido (3,11-14). Há também técnicas de tamanho de amostra para outros delineamentos, como estudos de fatores de risco genéticos ou de potenciais genes candidatos (15-17), estudos econômicos (18-20), estudos de dose-resposta (21) ou estudos que envolvam mais de dois grupos (22). A internet é uma ótima fonte para esses procedimentos mais elaborados: por exemplo, faça uma busca por *sample size* (“tamanho de amostra”) e *logistic regression* (“regressão logística”).

No entanto, geralmente é mais fácil, pelo menos para investigadores iniciantes, estimar o tamanho de amostra supondo um método mais simples de análise, como o teste do qui-quadrado ou o teste *t*. Utilizar esses métodos também é uma boa forma de verificar os resultados obtidos quando se utilizam métodos mais sofisticados. Suponha, por exemplo, que um investigador esteja planejando um estudo de caso-controle para saber se o nível de colesterol sérico (variável contínua) está associado à ocorrência de tumores cerebrais (variável dicotômica). Mesmo se o plano eventual for analisar os dados com regressão logística, uma estimativa grosseira do tamanho de amostra pode ser obtida a partir do teste *t*. Tem sido constatado que as abordagens simplificadas geralmente produzem estimativas de tamanho de amostra semelhantes às geradas por técnicas mais elaboradas. No entanto, se uma proposta de auxílio que envolve altos

custos for submetida para financiamento, deve-se consultar um estatístico experiente. Os revisores de propostas para financiamento esperam que você use uma abordagem mais sofisticada, mesmo se eles perceberem que as estimativas do tamanho de amostra baseiam-se em “chutes” sobre qual seria o risco do desfecho, a magnitude de efeito e assim por diante. Quando você pede auxílio a um estatístico para estimar o tamanho de amostra, isso também passa a mensagem de que você tem acesso aos colaboradores que serão necessários para gerenciar e analisar os dados do estudo. De fato, um bioestatístico irá contribuir de muitas outras formas para o delineamento e execução do estudo. Mesmo assim, o estatístico certamente irá apreciar trabalhar com um pesquisador clínico que pensou sobre essas questões e pelo menos fez uma tentativa inicial de estimativa do tamanho de amostra.

Ensaios clínicos de equivalência e de não inferioridade

Às vezes, a meta de um estudo é descartar uma associação substancial entre as variáveis preditoras e as de desfecho. Um **ensaio clínico de equivalência** testa se um novo medicamento tem praticamente a mesma eficácia que um medicamento já estabelecido. É um desafio planejar o tamanho de amostra em um estudo como esse, pois a magnitude de efeito desejada é zero ou muito pequena. Um **ensaio clínico de não inferioridade** é uma versão unilateral desse delineamento e examina se um novo medicamento é pelo menos não substancialmente inferior ao medicamento já estabelecido (Capítulo 11).

Os cálculos de tamanho de amostra para esses delineamentos são complexos (23-26), e o auxílio de um estatístico experiente é recomendável. Um método aceitável é delinear o estudo de forma que tenha poder substancial (p. ex., 0,90 ou 0,95) para rejeitar a hipótese nula quando a magnitude de efeito for tão pequena que não seja clinicamente importante (p. ex., uma diferença de 5 mg/dL na média da glicemia de jejum). Se os resultados de um estudo com grande poder forem “negativos” (p. ex., se o intervalo de confiança de 95% excluir a diferença pré-especificada de 5 mg/dL), pode-se concluir que os dois medicamentos são equivalentes. Um problema com os estudos de equivalência e de não inferioridade é que o poder adicional e a magnitude pequena de efeito exigem um grande tamanho de amostra; entre esses dois tipos de

delineamento, o de não inferioridade tem a vantagem de ser unilateral, permitindo um menor tamanho de amostra ou um alfa menor.

Outro problema é a perda das salvaguardas comuns contra erros tipo I (rejeitar falsamente a hipótese nula). Essas salvaguardas são inerentes ao paradigma da hipótese nula e protegem um estudo convencional que compara um medicamento ativo contra um placebo. O paradigma da hipótese nula assegura que problemas no delineamento ou na execução de um estudo, como uso de medidas imprecisas ou perda excessiva no seguimento, tornem mais difícil rejeitar a hipótese nula. Portanto, em um estudo convencional que tenta rejeitar uma hipótese nula, há um forte incentivo para que se faça o melhor estudo possível. Em um estudo de não inferioridade, entretanto, em que o objetivo é não encontrar uma diferença, essas salvaguardas não se aplicam.

TÉCNICAS DE TAMANHO DE AMOSTRA PARA ESTUDOS DESCRITIVOS

Princípios um tanto diferentes também norteiam a estimativa de tamanho de amostra para estudos descritivos, incluindo os estudos sobre testes diagnósticos. Tais estudos não têm variáveis preditora e de desfecho, tampouco comparam estatisticamente grupos diferentes. Portanto, os conceitos de poder e de hipóteses nula e alternativa não se aplicam. Nesses estudos, o investigador calcula estatísticas descritivas, como médias e proporções. Mesmo assim, é comum estudos descritivos (*Qual é a prevalência de depressão em pacientes idosos em um ambulatório*?) fazerem perguntas analíticas (*Quais são os preditores de depressão nesses pacientes*?). Nesse caso, o cálculo do tamanho de amostra também deve considerar o estudo analítico, para evitar um problema comum que é obter poder insuficiente para o que é uma questão de interesse maior.

Estudos descritivos costumam relatar **intervalos de confiança**, ou seja, uma faixa de valores para a média ou proporção da amostra. O intervalo de confiança é uma medida de precisão de uma estimativa amostral. O investigador determina o nível de confiança, como, por exemplo, 95% ou 99%. Um intervalo com nível de confiança maior (p. ex., 99%), por ser mais amplo, tem maior probabilidade de incluir o valor populacional real do que um intervalo com um nível de confiança menor (p. ex., 90%).

A amplitude de um intervalo de confiança depende do tamanho de amostra. Por exemplo, um investigador pode querer estimar o escore médio do Exame de Certificação Médica dos Estados Unidos (U.S. Medical Licensing Examination) em um grupo de estudantes de medicina que estudaram usando um currículo alternativo pela internet. A partir de uma amostra de 50 estudantes, ele poderia estimar o escore médio na população de todos os estudantes como de 215, com um intervalo de confiança de 95% de 205 a 225. Um estudo menor, digamos com 20 estudantes, pode ter praticamente o mesmo escore médio, mas é quase certo que teria um intervalo de confiança de 95% mais amplo.

Na estimativa do tamanho de amostra para estudos descritivos, especifica-se o nível e a amplitude desejados para o intervalo de confiança. O tamanho de amostra pode, então, ser determinado a partir das tabelas ou fórmulas contidas no apêndice.

Variáveis contínuas

Para variáveis contínuas, é comum relatar um intervalo de confiança em torno do valor médio da variável. A estimativa do tamanho de amostra para esse intervalo de confiança (Exemplo 6.4) apresenta os seguintes passos:

1. Estimar o desvio-padrão da variável de interesse.
2. Especificar a precisão desejada (amplitude total) do intervalo de confiança.
3. Selecionar o nível de confiança para o intervalo (p. ex., 95%, 99%).

EXEMPLO 6.4 Cálculo do tamanho de amostra para um estudo descritivo de uma variável contínua

Problema: O investigador deseja determinar a média do nível de hemoglobina em alunos da terceira série em uma região urbana com um intervalo de confiança de 95% de ± 0,3 g/dL. Um estudo anterior apontou que o desvio-padrão da hemoglobina em uma cidade semelhante foi de 1 g/dL.

Solução: Os ingredientes para o cálculo do tamanho de amostra são apresentados a seguir:

1. Desvio-padrão da variável (DP) = 1 g/dL.

2. Amplitude total do intervalo = 0,6 g/dL (0,3 g/dL acima e 0,3 g/dL abaixo). Portanto, a amplitude padronizada do intervalo = amplitude total ÷ S = 0,6 ÷ 1 = 0,6.
3. Nível de confiançα = 95%.

Partindo de uma amplitude padronizada de 0,6 na coluna da esquerda da Tabela 6D e de um intervalo de confiança de 95%, o tamanho de amostra exigido é de 43 alunos de terceira série.

Para usar o Apêndice 6D, é preciso padronizar a amplitude total do intervalo (dividi-la pelo desvio-padrão da variável) e, então, localizar, na coluna da esquerda da Tabela 6D, a amplitude padronizada esperada. Feito isso, basta localizar horizontalmente o nível de confiança para obter o tamanho de amostra correspondente.

Variáveis dicotômicas

Em um estudo descritivo com uma variável dicotômica, os resultados podem ser expressos como um intervalo de confiança em torno da proporção estimada de sujeitos com um dos valores. Isso inclui estudos sobre a **sensibilidade** ou a **especificidade** de um teste diagnóstico, que, à primeira vista, parecem ser variáveis contínuas, mas são na verdade dicotômicas – proporções expressas como percentuais (Capítulo 12). Para estimar o tamanho de amostra para esse intervalo de confiança, deve-se:

1. Estimar a proporção esperada com a característica de interesse na população. (Caso se espere que mais da metade da população tenha a característica, deve-se planejar o tamanho de amostra com base na proporção que se acredite não ter a característica.)
2. Especificar a precisão desejada (amplitude total) para o intervalo de confiança.
3. Selecionar o nível de confiança para o intervalo (p. ex., 95%).

No Apêndice 6E, localize na coluna da esquerda da Tabela 6E a proporção esperada com a característica de interesse. Então, localize horizontalmente, a partir da amplitude e do intervalo de confiança escolhidos, o tamanho de amostra necessário.

O Exemplo 6.5 mostra o cálculo do tamanho de amostra para estimar a sensibilidade de um teste diagnóstico, indicando o número necessário de

sujeitos com a doença. Ao estudar a especificidade do teste diagnóstico, o investigador deve estimar o tamanho de amostra de sujeitos que não apresentam a doença em questão. Há também técnicas para estimar o tamanho de amostra para estudos que usam como desfecho curvas ROC (do inglês, *receiver operating characteristic*) (27), razões de verossimilhança (28) e confiabilidade (29) (Capítulo 12).

EXEMPLO 6.5 Cálculo do tamanho de amostra para um estudo descritivo com uma variável dicotômica

Problema: O investigador quer determinar a sensibilidade de um novo exame diagnóstico para câncer de pâncreas. Com base em um estudo-piloto, espera que 80% dos pacientes com câncer de pâncreas tenham resultados positivos. Quantos pacientes com esse tipo de câncer serão necessários para estimar uma sensibilidade de 0,80 ± 0,05 e um intervalo de confiança de 95% para esse novo exame?

Solução: Os ingredientes para o cálculo do tamanho de amostra são apresentados a seguir:

1. Proporção esperadα = 0,20. (Como 0,80 é mais da metade, o tamanho de amostra passa a ser estimado a partir da proporção que se espera que tenha resultado falsamente negativo, isto é, 0,20.)
2. Amplitude total = 0,10 (0,05 abaixo e 0,05 acima).
3. Nível de confiançα = 95%.

Partindo de 0,20 na coluna da esquerda da Tabela 6E e indo para baixo até uma amplitude total de 0,10, o número do meio (representando um nível de confiança de 95%) indica o tamanho de amostra exigido: 246 pacientes com câncer de pâncreas.

O QUE FAZER QUANDO O TAMANHO DE AMOSTRA É FIXO

Há casos em que o tamanho de amostra já está determinado antes do planejamento do estudo, especialmente quando se faz análise de dados secundários. Mesmo quando você está delineando um estudo em que os dados αinda não foram coletados, é comum descobrir que o número de participantes disponíveis ou acessíveis para o estudo é limitado. De fato, a

maioria dos investigadores, se forem honestos, irão admitir que frequentemente trabalham “de trás para frente”, a partir de um tamanho de amostra fixo ou realista, para determinar a magnitude de efeito que eles terão um poder razoável para detectar. Isso explica em parte porque não faz sentido tratar uma estimativa de tamanho de amostra como se fosse um número talhado em uma pedra.

Quando um investigador precisa trabalhar na direção oposta, isto é, partindo do tamanho de amostra, ele estima a magnitude de efeito que poderá ser detectada com um determinado poder estatístico (geralmente 80%) ou, o que é menos comum, o poder para detectar um determinado efeito. É possível, então, usar as tabelas de tamanho de amostra contidas nos apêndices deste capítulo, interpolando quando necessário, ou usar as fórmulas de tamanho de amostra dos apêndices para estimar a magnitude de efeito.

Uma regra geral é que um estudo deve ter um poder de 80% ou mais para detectar uma magnitude de efeito razoável. No entanto, não há nada de mágico em relação ao valor de 80%. Às vezes, um investigador pode ter sorte e achar um resultado estatisticamente significativo mesmo quando tiver poder limitado para isso (mesmo um poder estatístico tão baixo quanto 50% fornece uma probabilidade de 50% de observar um efeito estatisticamente significativo na amostra que está realmente presente na população). Portanto, pode valer a pena fazer estudos com poder estatístico menor do que 80% se o custo associado for baixo, por exemplo quando se analisam dados que já foram coletados. E há alguns estudos – por exemplo, um que mostra que um novo tratamento reduz a pressão arterial pulmonar em mais de 50% em pacientes com hipertensão pulmonar refratária – nos quais um tamanho de amostra de dois ou três sujeitos será suficiente para indicar que vale a pena realizar novos estudos sobre o tema, em relação à segurança e aos efeitos em desfechos clínicos.

No entanto, é importante que o investigador tenha em mente que poderá futuramente enfrentar dificuldades para interpretar (e publicar) um estudo que não encontrou uma associação por falta de poder estatístico; intervalos de confiança amplos irão mostrar a possibilidade de um efeito substancial na população da qual a amostra pequena foi selecionada. Também é importante compreender que um estudo com baixo poder e que teve sorte de encontrar um resultado estatisticamente significativo pode

ser criticado, pois os revisores podem ficar na dúvida sobre se o investigador realmente queria procurar aquela associação ou se simplesmente ficou testando diferentes hipóteses e selecionou o resultado que tinha um valor *P* estatisticamente significativo.

EXEMPLO 6.6 Cálculo da magnitude de efeito detectável quando o tamanho de amostra é fixo

Problema: Uma investigadora estima que terá acesso a 200 puérperas de gêmeos durante seu *fellowship*. Com base em um pequeno estudo-piloto, ela estima que em torno de metade dessas mulheres (isto é, 100) poderão estar dispostas a participar de um ensaio clínico sobre o efeito de um programa de meditação de seis semanas sobre o estresse. O grupo-controle receberá um folheto informativo orientando a técnica de relaxamento. Se o desvio-padrão esperado para o escore do estresse for de 5 pontos, tanto para o grupo experimental como para o grupo-controle, qual a magnitude da diferença que o investigador poderá detectar entre os dois grupos, para um α (bilateral) = 0,05 e β = 0,20?

Solução: Na Tabela 6A, partindo de α (bilateral) = 0,05 e β = 0,20 (coluna da direita na tríade de números do meio), são necessários 45 pacientes por grupo para detectar uma magnitude padronizada de efeito de 0,6, resultando em um total de 3 pontos (0,6 × 5 pontos). Portanto, o investigador (que terá aproximadamente 50 pacientes por grupo) poderá detectar uma diferença de um pouco menos de 3 pontos entre os dois grupos.

ESTRATÉGIAS PARA MINIMIZAR O TAMANHO DE AMOSTRA E MAXIMIZAR O PODER

Quando o tamanho de amostra estimado é maior do que o número de sujeitos que podem ser estudados de forma realista, deve-se proceder da seguinte forma. Primeiro os cálculos devem ser verificados, pois é fácil cometer erros. Então os “ingredientes” devem ser revisados. A magnitude de efeito é pequena demais, ou a variabilidade grande demais? O α e o β. foram estipulados em um valor razoável? O nível de confiança é alto

demais, ou o intervalo de confiança, muito estreito?

Esses ajustes técnicos podem ser úteis, mas é importante perceber que os testes estatísticos dependem das informações contidas nos dados. Muitas mudanças nos ingredientes, como a redução do poder de 90 para 80%, não alteram a quantidade ou a qualidade dos dados que serão coletados. Há, no entanto, várias estratégias para reduzir o tamanho de amostra necessário ou aumentar o poder para um determinado tamanho de amostra que realmente aumentam o conteúdo informacional dos dados coletados. Muitas dessas estratégias envolvem mudanças na hipótese de pesquisa; o investigador deve avaliar com cuidado se a nova hipótese ainda responde à questão de pesquisa que ele pretendia estudar.

Usar variáveis contínuas

Quando é possível usar variáveis contínuas, elas costumam possibilitar tamanhos de amostra menores do que as variáveis dicotômicas. A pressão arterial, por exemplo, pode ser expressa em milímetros de mercúrio (contínua) ou como a presença ou ausência de hipertensão (dicotômica). A primeira permite um tamanho de amostra menor para um determinado poder ou um poder maior para um determinado tamanho de amostra.

No Exemplo 6.7, o desfecho contínuo expressa o efeito de suplementos alimentares sobre a força muscular em idosos. O desfecho dicotômico expressa os efeitos dos suplementos sobre a proporção de idosos com força muscular acima de um valor mínimo, o que o torna um desfecho substituto mais apropriado para morbidades relacionadas com quedas.

EXEMPLO 6.7 O uso de variáveis contínuas *versus* dicotômicas

Problema: Considere um ensaio clínico randomizado controlado por placebo para determinar o efeito de suplementos alimentares na força muscular em idosos que moram em asilos. Estudos anteriores revelaram que a força muscular do quadríceps (pico de torque, em Newton-metros- Nm) tem uma distribuição aproximadamente normal, com uma média de 33 Nm e um desvio-padrão de 10 Nm, e que aproximadamente 10% dos idosos têm músculos muito fracos (força , 20 Nm). Sabe-se que o consumo de suplementos alimentares durante seis meses, comparado à dieta habitual, aumenta em 5 Nm a força

muscular. Com base na distribuição da força do quadríceps em idosos, estima-se que essa mudança na força média corresponda a uma redução de 5% na proporção de idosos com músculos muito fracos.

Uma opção seria tratar a força muscular como uma variável dicotômica, isto é, muito fraco *versus* não muito fraco. Outra opção seria usar todas as informações contidas na medida e tratar a força muscular como variável contínua. Quantos idosos cada delineamento exigiria para α (bilateral) = 0,05 e β = 0,20?

Solução: Os ingredientes para o cálculo de tamanho de amostra usando-se uma **variável de desfecho dicotômica** (muito fraco ou não muito fraco) são os seguintes:

1. **Hipótese nula:** a proporção de idosos residentes em asilos que são muito fracos (pico de torque do quadríceps < 20 Nm) e receberam suplemento alimentar durante seis meses é a mesma que aquela em idosos que são muito fracos e receberam dieta habitual.

   Hipótese alternativa: a proporção de idosos residentes em asilos que são muito fracos (pico de torque do quadríceps < 20 Nm) e receberam suplemento alimentar durante seis meses é diferente daquela em idosos que receberam dieta habitual
2. P_1 (proporção de idosos com dieta habitual que são muito fracos) = 0,10; P_2 (proporção de idosos que recebem o suplemento alimentar e que são muito fracos) = 0,05. O menor desses valores é 0,05, e a diferença entre eles ($P_1 - P_2$) é 0,05.
3. α (bilateral) = 0,05; β = 0,20.

Na Tabela 6B.1, partindo de 0,05 na coluna da esquerda e de uma diferença esperada de 0,05, constata-se pelo número do meio (α [bilateral] = 0,05 e β = 0,20) que o delineamento exigiria 473 idosos em cada grupo.

Os ingredientes para o cálculo do tamanho de amostra, usando-se uma **variável de desfecho contínua** (força do quadríceps como pico de torque), são os seguintes:

1. **Hipótese nula:** a força média do quadríceps (pico de torque em Nm) em idosos que residem em asilos e receberam suplementos

alimentares durante seis meses é a mesma que aquela em idosos que receberam dieta habitual.

Hipótese alternativa: a força média do quadríceps (pico de torque em Nm) em idosos que residem em asilos e receberam suplementos alimentares durante seis meses é diferente daquela em idosos que receberam dieta habitual.

2. Magnitude de efeito = 5 Nm.
3. Desvio-padrão da força muscular do quadríceps = 10 Nm.
4. Magnitude padronizada de efeito = magnitude do efeito ÷ desvio-padrão = 5 Nm ÷ 10 Nm = 0,5.
5. α (bilateral) = 0,05; β = 0,20.

Usando-se a Tabela 6A, partindo de uma magnitude padronizada de efeito de 0,50, para α (bilateral) = 0,05 e β = 0,20, esse delineamento exigiria aproximadamente 64 idosos em cada grupo. (Neste exemplo, a fórmula simplificada para o cálculo do tamanho de amostra apresentada na seção Técnicas de tamanho de amostra para estudos analíticos e experimentais, tópico Teste *t* – Tamanho da amostra (por grupo de igual tamanho) = 16 ÷ (magnitude padronizada de efeito)^2 – gera o mesmo tamanho de amostra: 16 ÷ (magnitude padronizada de efeito)^2 = 16 ÷ (0,5)^2 = 64 idosos por grupo.) Em resumo, o importante é notar que o uso de uma variável de desfecho contínua resultou em um tamanho de amostra muito menor.

Usar medidas pareadas

Em certos ensaios clínicos ou estudos de coorte com variáveis de desfecho contínuas, é possível fazer medições pareadas para cada sujeito, uma na linha de base e outra na conclusão do estudo. A variável de desfecho passa a ser a mudança entre essas duas medidas. Nesse caso, pode-se usar um teste *t* para medidas pareadas comparando-se o valor médio da mudança entre os dois grupos. Essa técnica pode permitir um tamanho de amostra menor, pois, ao comparar cada sujeito com ele próprio, acaba-se com a parte "intersujeito" da variabilidade da variável de desfecho. Por exemplo, a mudança de peso em uma dieta tem uma variabilidade menor do que o peso final porque o peso final é altamente

correlacionado com o peso inicial. O tamanho de amostra para esse tipo de teste *t* é estimado como na forma usual (Exemplo 6.8); a única diferença é que a magnitude padronizada de efeito (*E/DP* na Tabela 6A) é a diferença antecipada da *mudança* da variável dividida pelo desvio-padrão *dessa mudança*.

Breve comentário técnico

Neste capítulo, sempre se fala sobre **testes *t* para duas amostras**, testes esses usados para comparar valores médios de variáveis contínuas em dois grupos de sujeitos. Um teste *t* para duas amostras pode ser **não pareado**, se a variável estiver sendo comparada em dois grupos (Exemplo 6.1), ou **pareado**, se a variável for a mudança em um par de medidas (p. ex., antes e depois de uma intervenção (veja Exemplo 6.8).

EXEMPLO 6.8 Teste *t* em medidas pareadas

Problema: Retome o Exemplo 6.1 sobre o tratamento da asma, em que o investigador testa se o salbutamol, comparado ao brometo de ipratrópio, produz um aumento adicional de 200 mL no VEF_1. Os cálculos de tamanho de amostra indicaram a necessidade de 394 pacientes por grupo, um número maior que o disponível. Felizmente, um colaborador lembra que pacientes asmáticos apresentam grande variação nos valores basais de VEF_1. Essas diferenças interindivíduos são responsáveis por grande parte da variabilidade do VEF_1 após o tratamento, obscurecendo, assim, a medida do efeito terapêutico. Ele, então, propõe usar um teste *t* pareado (para duas amostras) para comparar as mudanças do VEF_1 nos dois grupos. Um estudo-piloto mostra que o desvio-padrão da mudança no VEF_1 é de apenas 250 mL. Quantos sujeitos seriam necessários por grupo, para α (bilateral) = 0,05 e β = 0,20?

Solução: Os ingredientes para o cálculo do tamanho de amostra são mostrados a seguir:

1. **Hipótese nula:** a mudança no VEF_1 médio duas semanas após o tratamento é a mesma em pacientes asmáticos tratados com salbutamol e nos tratados com brometo de ipratrópio.

Hipótese alternativa: a mudança no VEF_1 médio duas semanas após o tratamento em pacientes asmáticos tratados com salbutamol é diferente daquela em pacientes tratados com brometo de ipratrópio.

2. Magnitude de efeito = 200 mL.
3. Desvio-padrão da variável de desfecho = 250 mL.
4. Magnitude padronizada de efeito = magnitude do efeito ÷ desvio-padrão = 200 mL ÷ 250 mL = 0,8.
5. α (bilateral) = 0,05; $\beta = 1 - 0{,}80 = 0{,}20$.

De acordo com a Tabela 6A, esse delineamento exigiria, aproximadamente, 26 participantes por grupo, um tamanho de amostra bem mais razoável que o de 394 por grupo do Exemplo 6.1. Neste exemplo, a fórmula simplificada para o cálculo do tamanho de amostra produz uma estimativa bastante similar: 16 ÷ (magnitude padronizada de efeito)^2 = 16 ÷ $(0{,}8)^2$ = 25 sujeitos por grupo.

Um terceiro tipo de teste *t*, **o teste *t* para uma única amostra pareada**, compara a mudança média em um par de valores em um único grupo com uma mudança de zero. Esse tipo de análise é relativamente comum em delineamentos de séries temporais (Capítulo 11), uma abordagem antes-depois para examinar tratamentos que são difíceis de randomizar (p. ex., o efeito de histerectomia eletiva na qualidade de vida, uma decisão que poucas mulheres estariam dispostas a permitir que fosse tomada com base no resultado do lançamento de uma moeda). Entretanto, é um delineamento mais fraco, porque a ausência de um grupo de comparação torna difícil saber o que poderia ter acontecido se os pacientes não tivessem sido tratados. Ao planejar um estudo que será analisado mediante um teste *t* para uma única amostra pareada, o tamanho de amostra total é apenas metade do tamanho de amostra por grupo listado no Apêndice 6A. Por exemplo, para α (bilateral) = 0,05 e β. = 0,20, para detectar uma diferença do desvio-padrão de 0,05 (E/DP = 0,05) seriam necessários 64/2 = 32 sujeitos. O Apêndice 6F apresenta informações adicionais sobre o uso e o abuso dos testes *t* para uma e duas amostras.

Usar variáveis mais precisas

Por reduzirem a variabilidade, variáveis mais precisas possibilitam uma diminuição do tamanho de amostra necessário para estudos analíticos e descritivos. Mesmo uma mudança pequena na precisão pode ter um efeito significativo. Por exemplo, na estimativa do tamanho de amostra com um teste *t*, um decréscimo de 20% no desvio-padrão da variável de desfecho resultaria em um decréscimo de 36% no tamanho da amostra. Técnicas para aumentar a precisão de uma variável, tais como medidas em duplicata, são apresentadas no Capítulo 4.

Usar tamanhos desiguais para os grupos

Estudos com número igual de sujeitos nos dois grupos geralmente alcançam o maior poder estatístico para um dado número total de sujeitos. Dessa forma, as Tabelas 6A, 6B.1 e 6B.2 nos apêndices pressupõem tamanhos iguais de amostra nos dois grupos. Às vezes, no entanto, a distribuição de sujeitos não é igual nos dois grupos, ou é mais fácil ou mais barato recrutar sujeitos para um grupo que para o outro. Pode acontecer, por exemplo, de um investigador querer estimar o tamanho de amostra para um estudo comparando os 30% dos sujeitos de uma coorte que fumam com os não 70% que não fumam. Outro exemplo seria um estudo de caso-controle em que o número de indivíduos com a doença é pequeno, mas em que é possível selecionar um número muito maior de controles. Em geral, há um ganho considerável de poder quando o tamanho de um grupo é o dobro do tamanho do outro; no entanto, triplicar e quadruplicar o tamanho de um dos grupos fornece ganhos progressivamente menores. Os tamanhos de amostra para grupos desiguais podem ser computados a partir das fórmulas encontradas no texto dos Apêndices 6A e 6B ou em calculadores de tamanho de amostra em softwares estatísticos ou na internet.

A seguir, apresentamos uma aproximação útil (30) para estimar o tamanho de amostra para estudos de caso-controle sobre fatores de risco e desfechos dicotômicos, usando *c* controles por caso (Exemplo 6.9). Se *n* representa o número de casos necessários quando há 1 controle para cada caso (para α, β. e magnitude de efeito especificados), então o número aproximado de casos (*n*') necessários quando há *cn*' controles é:

$$n' = [(c + 1)] \div 2c\ ] \times n$$

EXEMPLO 6.9 Uso de múltiplos controles por caso em um estudo de caso-controle

Problema: Um investigador está estudando se a exposição a inseticidas no ambiente doméstico é fator de risco para anemia aplástica. O cálculo de tamanho de amostra original indicou que seriam necessários 25 casos, considerando-se um controle por caso. Suponha que o investigador tenha acesso a apenas 18 casos. Então, como deveria proceder?

Solução: O investigador deveria considerar o uso de múltiplos controles por caso (afinal, ele pode encontrar vários pacientes sem anemia aplástica). Usando três controles por caso, por exemplo, o número aproximado de casos necessários é $[(3 + 1) \div (2 \times 3)] \times 25 = 17$.

Por exemplo, com $c = 2$ controles por caso, $[(2 + 1) \div (2 \times 2)] \times n = 3/4 \times n$, isso significa que são necessários apenas 75% dos casos. À medida que c aumenta, n' se aproxima de 50% de n (p. ex., quando $c = 10$, $n' = 11/20 \times n$).

Usar um desfecho mais comum

Ao planejar um estudo com um desfecho dicotômico, quanto mais frequente for o desfecho, até uma frequência máxima de aproximadamente 0,5, maior o poder estatístico. Portanto, mudar a definição de um desfecho é uma das melhores formas de aumentar o poder. O poder, na verdade, depende mais do número de sujeitos que apresentam um determinado desfecho do que do número total de sujeitos no estudo. Estudos com desfechos raros, como a ocorrência de câncer de mama em mulheres hígidas, exigem tamanhos de amostra muito grandes para alcançar poder estatístico adequado.

Uma das melhores formas de tornar um desfecho mais comum é arrolar sujeitos com um risco maior de desenvolver o desfecho (p. ex., mulheres com história familiar de câncer de mama). Outras formas são estender o período de seguimento, permitindo maior acúmulo de desfechos, ou simplificar a definição do que constitui um desfecho (p. ex., incluir carcinoma ductal *in situ*). É importante ressaltar que essas técnicas podem

alterar a questão de pesquisa e devem ser usadas com cuidado (Exemplo 6.10).

EXEMPLO 6.10 Uso de um desfecho mais comum

Problema: Imagine que uma investigadora esteja comparando a eficácia de uma solução antisséptica para gargarejo com a de uma solução-placebo na prevenção de infecções de vias aéreas superiores. Os cálculos iniciais indicaram que a amostra que ela havia planejado, de 200 estudantes universitários voluntários, era inadequada, em parte, porque previa que apenas 20% dos sujeitos teriam infecção de vias aéreas superiores durante o período de seguimento de três meses. Sugira algumas mudanças no plano de estudo.

Solução: Três soluções possíveis são: (1) estudar uma amostra de residentes de pediatria, pois provavelmente apresentam incidência muito maior de infecções de vias aéreas superiores que os estudantes universitários; ou (2) realizar o estudo no inverno, quando essas infecções são mais comuns; ou (3) seguir a amostra por um período mais longo, como 6 ou 12 meses. Todas essas soluções envolvem mudanças na hipótese de pesquisa, mas sem afetar de forma significativa a questão de pesquisa em termos da eficácia do gargarejo de solução antisséptica.

COMO ESTIMAR O TAMANHO DE AMOSTRA QUANDO AS INFORMAÇÕES SÃO INSUFICIENTES

Muitas vezes, o investigador descobre que faltam um ou mais ingredientes para o cálculo do tamanho de amostra e se frustra ao tentar planejar o estudo. Esse problema é bastante comum quando se usa um instrumento desenvolvido especificamente para essa pesquisa (p. ex., um novo questionário que compara a qualidade de vida em mulheres com incontinência urinária de estresse *versus* de urgência). Como decidir que fração de um desvio-padrão dos escores obtidos pelo instrumento seria clinicamente relevante?

A primeira estratégia é fazer uma **busca extensa** sobre estudos anteriores relacionados a esse assunto e questões de pesquisa semelhantes.

Situações apenas grosseiramente comparáveis e achados de baixa qualidade ou desatualizados já podem ser úteis. Por exemplo, existem dados sobre qualidade de vida em pacientes com outros problemas urológicos ou com condições de saúde relacionadas, como vivendo com uma colostomia? Se a revisão da literatura não for produtiva, deve-se então consultar outros investigadores sobre o que esperar, ou sobre a existência de algum estudo não publicado que possa ser relevante.

Se ainda faltarem informações, o investigador pode considerar obter os ingredientes que faltam por meio de um pequeno **estudo-piloto** ou da análise de dados secundários, antes de se aventurar no estudo principal. Na verdade, um estudo-piloto é fortemente recomendado para praticamente todos os estudos que envolvem novos instrumentos, métodos de mensuração ou estratégias de recrutamento. Ele economiza tempo, pois permite que o pesquisador planeje o estudo principal de maneira muito melhor. Estudos-piloto permitem estimar o desvio-padrão de uma medida ou a proporção de sujeitos com uma característica específica. Entretanto, uma alternativa é reconhecer que, no caso de variáveis contínuas com distribuições próximas da normalidade, o **desvio-padrão** pode ser estimado em um quarto da amplitude entre os limites superior e inferior de uma faixa comum de valores, ignorando valores extremos. Por exemplo, se a maioria dos sujeitos tem grande probabilidade de apresentar um nível de sódio sérico entre 135 e 143 mEq/L, o desvio-padrão do sódio sérico seria estimado em torno de 2 mEq/L (1/4 × 8 mEq/L).

Outra estratégia, quando há dúvidas sobre a média e o desvio-padrão de uma variável contínua ou categórica, é **dicotomizar** essa variável. As variáveis categóricas podem ser redistribuídas em dois grupos, e as contínuas podem ser divididas na sua média ou mediana. Por exemplo, dividir a qualidade de vida em “melhor do que a mediana” ou “igual ou pior do que a mediana” evita a necessidade de estimar o desvio-padrão na amostra, embora ainda seja necessário estimar a proporção de sujeitos com valores acima da mediana nos dois grupos sob investigação. O teste do qui-quadrado pode, então, ser usado para fazer uma estimativa razoável, embora levemente elevada, do tamanho de amostra.

Com frequência, no entanto, o investigador deve escolher a magnitude de efeito detectável com base em um valor que considere **clinicamente relevante**. Nesse caso, o investigador deve buscar aconselhamento a

respeito de sua escolha com colegas que trabalham nessa área. Por exemplo, imagine que um investigador está estudando um novo tratamento invasivo para gastroparesia grave refratária, uma doença na qual no máximo 5% dos pacientes melhoram espontaneamente. Seus colegas gastrenterologistas disseram que, se o tratamento se mostrar efetivo, estariam dispostos a tratar até 5 pacientes para produzir um benefício sustentado em apenas um deles (como o tratamento possui efeitos colaterais significativos e é caro, eles acham que não estariam dispostos a um número maior do que 5). Um número necessário tratar (NNT) de 5 corresponde a uma diferença de risco de 20% (NNT = 1/diferença de risco) e, portanto, o investigador deve estimar o tamanho de amostra baseando-se na comparação P_1 = 5% *versus* P_2 = 25% (isto é, 59 pacientes por grupo, considerando poder igual a 0,80 e α bilateral de 0,05).

Se tudo isso falhar, o investigador deve fazer um **"chute científico"** sobre os valores prováveis dos ingredientes que faltam. O processo de pensar o problema por etapas e imaginar os achados geralmente leva a uma estimativa razoável, e é isso que significa planejar o tamanho de amostra. Normalmente essa é uma opção melhor do que simplesmente decidir, na ausência de qualquer justificativa, delinear um estudo com poder de 80% e α bilateral de 0,05 para detectar uma magnitude padronizada de efeito de, digamos, 0,5 entre os dois grupos (n = 64 sujeitos por grupo, nesse caso). Raros revisores de projetos de pesquisa irão aceitar uma decisão inteiramente arbitrária como essa.

ERROS COMUNS QUE DEVEM SER EVITADOS

Muitos investigadores com pouca experiência (e até mesmo alguns com muita experiência!) cometem erros ao planejar o tamanho de amostra. Alguns dos erros mais comuns são:

1. Um erro comum é estimar o tamanho de amostra tarde demais no delineamento do estudo. É importante fazê-lo cedo no processo, quando mudanças fundamentais ainda podem ser feitas.
2. Variáveis dicotômicas podem aparentar ser contínuas quando expressas como **percentagem** ou **taxa**. Por exemplo, o estado vital

(vivo ou morto) pode ser interpretado erroneamente como contínuo quando expresso em termos da percentagem de indivíduos vivos. Da mesma forma, em uma análise de sobrevivência em que nem todos os sujeitos morrem, um desfecho dicotômico pode aparentar ser contínuo (p. ex., mediana da sobrevida em meses). Em todos esses casos, o desfecho é, na verdade, dicotômico (uma proporção), e a abordagem simples apropriada para o planejamento do tamanho de amostra seria o teste do qui-quadrado.

3. O tamanho de amostra estima o número de sujeitos que deverão apresentar dados sobre o desfecho, e não simplesmente o número que deverá ser incluído no estudo. O investigador deve sempre estar preparado para casos de **abandono (*dropouts*)** e para casos com **dados faltantes (*missing data*)**.
4. As tabelas no final deste capítulo pressupõem que os dois grupos em estudo têm tamanhos iguais de amostra. Esse nem sempre é o caso. Por exemplo, um estudo de coorte para saber se suplementos vitamínicos reduzem o risco de queimaduras solares provavelmente não poderia incluir número igual de sujeitos que usaram e que não usaram vitaminas. Se os grupos **não forem de mesmo tamanho**, deve-se usar as fórmulas que aparecem depois das tabelas ou calculadoras disponíveis na internet ou em softwares estatísticos.
5. Ao usar o teste *t* para estimar o tamanho de amostra, o desvio-padrão da variável de desfecho é um fator-chave. Portanto, se o desfecho for a **mudança** em uma variável contínua, o investigador deve usar o desvio-padrão da mudança, em vez do desvio-padrão da própria variável.
6. Cuidado com dados em **conglomerados**. Quando parecem existir dois "níveis" de tamanho de amostra (p. ex., um para médicos e outro para pacientes), os conglomerados podem ser um problema e as tabelas nos apêndices não são adequadas.
7. Se você tiver dificuldade para estimar um tamanho de amostra para seu estudo, assegure-se de que sua hipótese de pesquisa atende aos critérios discutidos no capítulo anterior (simples, específica e formulada antes dos fatos).

RESUMO

1. Para estimar o tamanho de amostra para um **estudo analítico**, deve-se proceder da seguinte forma:
   (a) formular as **hipóteses nula** e **alternativa**, especificando o **número de lados**;
   (b) selecionar um **teste estatístico** para analisar os dados, com base nos **tipos de variáveis preditoras e de desfecho** (**teste do qui-quadrado** se ambas as variáveis forem dicotômicas, **teste *t*** se uma variável for dicotômica e a outra contínua, e **coeficiente de correlação** se ambas forem contínuas);
   (c) estimar a **magnitude de efeito** e sua **variabilidade, se necessário**; e
   (d) estabelecer os valores apropriados para α e β., com base na importância de se evitar **erros tipo I** e **tipo II**.
2. Outros itens a serem considerados no cálculo de tamanho de amostra para estudos analíticos incluem ajustes para potenciais **abandonos** e estratégias para lidar com **variáveis categóricas**, **análise de sobrevivência**, **amostras por conglomerados**, **ajuste multivariado** e abordagens estatísticas especiais para **ensaios clínicos de equivalência e de não inferioridade**.
3. A estimativa de tamanho de amostra para **estudos descritivos** sem hipóteses apresenta os seguintes passos: (a) estimar a **proporção** de sujeitos com um desfecho dicotômico ou o **desvio-padrão** de um desfecho contínuo; (b) especificar a **precisão** desejada (amplitude do intervalo de confiança); e (c) especificar o **nível de confiança** (p. ex., 95%).
4. Quando o tamanho de amostra é predeterminado, o investigador pode inverter os procedimentos e estimar a **magnitude de efeito** detectável ou, o que é menos comum, o **poder estatístico** do estudo.
5. As estratégias **para minimizar** o tamanho de amostra incluem usar **variáveis contínuas**, **medidas mais precisas**, **medidas pareadas** e **desfechos mais comuns**, assim como aumentar o número de controles por caso em estudos de caso-controle.
6. Quando parece não haver informações suficientes para estimar o tamanho de amostra, o investigador deve **revisar a literatura** em

áreas afins e consultar **colegas** para ajudarem a escolher uma **magnitude de efeito** clinicamente relevante.

7. Erros a serem evitados incluem estimar o tamanho de amostra **tarde demais**, interpretar erroneamente **proporções expressas como percentagens**, não levar em consideração **sujeitos e dados faltantes** e não abordar **dados em conglomerados e pareados** de forma apropriada.

APÊNDICE 6A

Tamanho de amostra necessário por grupo quando se usa o teste *t* para comparar médias de variáveis contínuas

TABELA 6A Tamanho de amostra por grupo para comparar duas médias

α UNILATERAL =	0,005			0,025			0,05		
α BILATERAL =	**0,01**			**0,05**			**0,10**		
E/Dp* β =	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**
0,10	3.565	2.978	2.338	2.600	2.103	1.571	2.166	1.714	1.23
0,15	1.586	1.325	1.040	1.157	935	699	963	762	55
0,20	893	746	586	651	527	394	542	429	31
0,25	572	478	376	417	338	253	347	275	19
0,30	398	333	262	290	235	176	242	191	13
0,40	225	188	148	164	133	100	136	108	7
0,50	145	121	96	105	86	64	88	70	5
0,60	101	85	67	74	60	45	61	49	3
0,70	75	63	50	55	44	34	45	36	2
0,80	58	49	39	42	34	26	35	28	2
0,90	46	39	32	34	27	21	28	22	1
1,00	38	32	26	27	23	17	23	18	14

* E/DP é a magnitude padronizada de efeito, computada como E (magnitude de efeito esperada) dividido por DP (desvio-padrão da variável de desfecho). Para estimar o tamanho de amostra, localize o valor da magnitude padronizada de efeito e os valores especificados de α e β para o tamanho de amostra necessário em cada grupo. Para um teste t para uma amostra, o tamanho de amostra total é metade do número listado.

CALCULANDO A VARIABILIDADE

A variabilidade é geralmente relatada como desvio-padrão (DP) ou erro-padrão da média (EPM). Para fins de cálculo de tamanho de amostra, o DP da variável é a medida mais utilizada. Felizmente, é fácil converter de uma medida para a outra: o DP é simplesmente o erro-padrão vezes a raiz quadrada de *N*, onde *N* é o número de sujeitos na amostra em que a média foi obtida. Suponha que um estudo tenha relatado que a perda de peso em 25 indivíduos com uma dieta baixa em fibras foi de 10 ± 2 kg (média ± EPM). O DP seria $2 \times \sqrt{25}$ =10 Kg.

FÓRMULA GERAL PARA OUTROS VALORES

A fórmula geral para outros valores de *E*, *DP*, α e β, ou para grupos de tamanho desigual, onde *E* e *DP* são os definidos anteriormente, é apresentada a seguir. Para:

z_α = desvio normalizado da distribuição normal para α (Se a hipótese alternativa for bilateral, z_α = 2,58 quando α = 0,01, $z\alpha$ = 1,96 quando α = 0,05, e z_α = 1,645 quando α = 0,10. Se a hipótese alternativa for unilateral, z_α = 1,645 quando α = 0,05.)

z_β = desvio normalizado da distribuição normal para β (z_β = 0,84 quando β = 0,20, e z_β = 1,282 quando β = 0,10)

q_1 = proporção de sujeitos no grupo 1

q_2 = proporção de sujeitos no grupo 2

N = número **total** de sujeitos necessários

Então:

$$N = [(1/q_1 + 1/q_2)\ DP^2\ (Z_\alpha + Z_\beta)^2] \div E^2.$$

Aqueles que não desejarem fazer manualmente os cálculos envolvidos nesta fórmula podem obter os resultados instantaneamente usando uma calculadora disponível no nosso *website* www.epibiostat.ucsf.edu/dcr/. (Como essa fórmula se baseia na aproximação da estatística *t* com a estatística z, ela subestimará levemente o tamanho de amostra quando *N* for menor do que aproximadamente 30. A Tabela 6A usa a estatística *t* para estimar o tamanho de amostra.)

APÊNDICE 6B

Tamanho de amostra necessário por grupo ao usar a estatística do qui-quadrado ou o teste Z para comparar proporções de variáveis dicotômicas

TABELA 6B.1 Tamanho de amostra *por grupo* para comparar duas proporções

O MENOR ENTRE P_1 E P_2*	NÚMERO SUPERIOR: α = 0,05 (UNILATERAL) OU α = 0,10 (BILATERAL); β = 0,20 NÚMERO DO MEIO: α = 0,025 (UNILATERAL) OU α = 0,05 (BILATERAL); β = 0,20 NÚMERO INFERIOR: α = 0,025 (UNILATERAL) OU α = 0,05 (BILATERAL); β = 0,10									
	DIFERENÇA ENTRE P_1 E P_2									
	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0,05	381	129	72	47	35	27	22	18	15	13
	473	159	88	59	43	33	26	22	18	16
	620	207	113	75	54	41	33	27	23	19
0,10	578	175	91	58	41	31	24	20	16	14
	724	219	112	72	51	37	29	24	20	17
	958	286	146	92	65	48	37	30	25	21
0,15	751	217	108	67	46	34	26	21	17	15
	944	270	133	82	57	41	32	26	21	18
	1.252	354	174	106	73	53	42	33	26	22
0,20	900	251	121	74	50	36	28	22	18	15
	1.133	313	151	91	62	44	34	27	22	18
	1.504	412	197	118	80	57	44	34	27	23
0,25	1.024	278	132	79	53	38	29	23	18	15
	1.289	348	165	98	66	47	35	28	22	18
	1.714	459	216	127	85	60	46	35	28	23
0,30	1.123	300	141	83	55	39	29	23	18	15
	1.415	376	175	103	68	48	36	28	22	18
	1.883	496	230	134	88	62	47	36	28	23
0,35	1.197	315	146	85	56	39	29	23	18	15
	1.509	395	182	106	69	48	36	28	22	18

	2.009	522	239	138	90	62	47	35	27	22
0,40	1.246	325	149	86	56	39	29	22	17	14
	1.572	407	186	107	69	48	35	27	21	17
	2.093	538	244	139	90	62	46	34	26	21
0,45	1.271	328	149	85	55	38	28	21	16	13
	1.603	411	186	106	68	47	34	26	20	16
	2.135	543	244	138	88	60	44	33	25	19
0,50	1.271	325	146	83	53	36	26	20	15	—
	1.603	407	182	103	66	44	32	24	18	—
	2.135	538	239	134	85	57	42	30	23	—
0,55	1.246	315	141	79	50	34	24	18	—	—
	1.572	395	175	98	62	41	29	22	—	—
	2.093	522	230	127	80	53	37	27	—	—
0,60	1.197	300	132	74	46	31	22	—	—	—
	1.509	376	165	91	57	37	26	—	—	—
	2.009	496	216	118	73	48	33	—	—	—
0,65	1.123	278	121	67	41	27	—	—	—	—
	1.415	348	151	82	51	33	—	—	—	—
	1.883	459	197	106	65	41	—	—	—	—
0,70	1.024	251	108	58	35	—	—	—	—	—
	1.289	313	133	72	43	—	—	—	—	—
	1.714	412	174	92	54	—	—	—	—	—
0,75	900	217	91	47	—	—	—	—	—	—
	1.133	270	112	59	—	—	—	—	—	—
	1.504	354	146	75	—	—	—	—	—	—
0,80	751	175	72	—	—	—	—	—	—	—
	944	219	88	—	—	—	—	—	—	—
	1.252	286	113	—	—	—	—	—	—	—
0,85	578	129	—	—	—	—	—	—	—	—
	724	159	—	—	—	—	—	—	—	—
	958	207	—	—	—	—	—	—	—	—
0,90	381	—	—	—	—	—	—	—	—	—
	473	—	—	—	—	—	—	—	—	—
	620	—	—	—	—	—	—	—	—	—

As estimativas unilaterais usam a estatística z.
* P_1 representa a proporção esperada de sujeitos com o desfecho em um grupo; P_2, a mesma proporção no outro grupo. (Em um estudo de caso-controle, P_1 representa a proporção de casos com a variável preditora; P_2, a proporção de controles com a variável preditora.) Para estimar o tamanho de amostra, localize o menor entre P_1 e P_2 e a diferença esperada entre P_1 e P_2. Os três valores na tabela representam o tamanho de amostra necessário em cada grupo para os valores especificados de α e β.
Detalhes adicionais para P_1 e P_2 entre 0,01 e 0,10 são fornecidos na Tabela 6B.2.

TABELA 6B.2 Tamanho de amostra por grupo para comparar duas proporções, a menor delas estando entre 0,01 e 0,10

	NÚMERO SUPERIOR: α = 0,05 (UNILATERAL) OU α = 0,10 (BILATERAL); β = 0,20 NÚMERO DO MEIO: α = 0,025 (UNILATERAL) OU α = 0,05 (BILATERAL); β = 0,20 NÚMERO INFERIOR: α = 0,025 (UNILATERAL) OU α = 0,05 (BILATERAL); β = 0,10									
O MENOR ENTRE P_1 E P_2	**DIFERENÇA ESPERADA ENTRE P_1 E P_2**									
	0,01	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**	**0,10**
0,01	2.019	700	396	271	204	162	134	114	98	87
	2.512	864	487	332	249	197	163	138	120	106
	3.300	1.125	631	428	320	254	209	178	154	135
0,02	3.205	994	526	343	249	193	157	131	113	97
	4.018	1.237	651	423	306	238	192	161	137	120
	5.320	1.625	852	550	397	307	248	207	177	154
0,03	4.367	1.283	653	414	294	224	179	148	126	109
	5.493	1.602	813	512	363	276	220	182	154	133
	7.296	2.114	1.067	671	474	359	286	236	199	172
0,04	5.505	1.564	777	482	337	254	201	165	139	119
	6.935	1.959	969	600	419	314	248	203	170	146
	9.230	2.593	1.277	788	548	410	323	264	221	189
0,05	6.616	1.838	898	549	380	283	222	181	151	129
	8.347	2.308	1.123	686	473	351	275	223	186	159
	11.123	3.061	1.482	902	620	460	360	291	242	206
0,06	7.703	2.107	1.016	615	422	312	243	197	163	139
	9.726	2.650	1.272	769	526	388	301	243	202	171
	12.973	3.518	1.684	1.014	691	508	395	318	263	223
0,07	8.765	2.369	1.131	680	463	340	263	212	175	148
	11.076	2.983	1.419	850	577	423	327	263	217	183
	14.780	3.965	1.880	1.123	760	555	429	343	283	239
0,08	9.803	2.627	1.244	743	502	367	282	227	187	158
	12.393	3.308	1.562	930	627	457	352	282	232	195

	16.546	4.401	2.072	1.229	827	602	463	369	303	255
0,09	10.816	2.877	1.354	804	541	393	302	241	198	167
	13.679	3.626	1.702	1.007	676	491	377	300	246	207
	18.270	4.827	2.259	1.333	893	647	495	393	322	270
0,10	11.804	3.121	1.461	863	578	419	320	255	209	175
	14.933	3.936	1.838	1.083	724	523	401	318	260	218
	19.952	5.242	2.441	1.434	957	690	527	417	341	285

As estimativas unilaterais usam a estatística z.

FÓRMULA GERAL PARA OUTROS VALORES

A fórmula geral para calcular o tamanho ***total*** de amostra (*N*) necessário para um estudo usando a estatística z, onde P_1 e P_2 são definidos anteriormente, é apresentada a seguir (veja Apêndice 6A para definições de z_α e z_β.). Para:

q_1 = proporção de sujeitos no grupo 1

q_2 = proporção de sujeitos no grupo 2

N = número *total* de sujeitos

$P = q_1 P_1 + q_2 P_2$

Então:

$$N = \frac{[Z_\alpha \sqrt{P(1-P)(1/q_1 + 1/q_2)} + Z_\beta \sqrt{P_1(1-P_1)(1/q_1) + P_2(1-P_2)(1/q_2)}]^2}{(P_1 - P_2)^2}$$

Aqueles que não desejarem fazer manualmente os cálculos envolvidos nesta fórmula podem obter os resultados instantaneamente usando uma calculadora disponível no nosso *website* www.epibiostat.ucsf.edu/dcr/. (Esta fórmula não inclui a correção de continuidade de Fleiss-Tytun-Ury e, portanto, subestima o tamanho de amostra necessário em torno de até 10%. As Tabelas 6B.1 e 6B.2 incluem essa correção de continuidade.)

APÊNDICE 6C

Tamanho total de amostra necessário quando se usa o coeficiente de correlação (*r*)

TABELA 6C Tamanho de amostra para determinar se um coeficiente de correlação é diferente de zero

α UNILATERAL =	0,005			0,025			0,05		
α BILATERAL =	**0,01**			**0,05**			**0,10**		
r* **β =**	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**	**0,05**	**0,10**	**0,20**
0,05	7.118	5.947	4.663	5.193	4.200	3.134	4.325	3.424	2.46
0,10	1.773	1.481	1.162	1.294	1.047	782	1.078	854	61
0,15	783	655	514	572	463	346	477	378	27
0,20	436	365	287	319	259	194	266	211	15
0,25	276	231	182	202	164	123	169	134	9
0,30	189	158	125	139	113	85	116	92	6
0,35	136	114	90	100	82	62	84	67	4
0,40	102	86	68	75	62	47	63	51	3
0,45	79	66	53	58	48	36	49	39	2
0,50	62	52	42	46	38	29	39	31	2
0,60	40	34	27	30	25	19	26	21	1
0,70	27	23	19	20	17	13	17	14	1
0,80	18	15	13	14	12	9	12	10	8

* Para estimar o tamanho total de amostra, partindo do r (coeficiente de correlação esperado), mova horizontalmente até a coluna dos valores especificados para α e β.

FÓRMULA GERAL PARA OUTROS VALORES

A fórmula geral para outros valores de *r*, α, e β é a seguinte (veja Apêndice 6A para definições de Z_α e Z_β). Para:

r = coeficiente de correlação esperado

$C = 0,5 \times \ln[(1 + r)/(1 - r)]$

N = Número total de sujeitos necessário

Então:

$$N = [(Z_\alpha + Z_\beta) \div C]^2 + 3.$$

ESTIMANDO O TAMANHO DE AMOSTRA PARA A DIFERENÇA ENTRE DUAS CORRELAÇÕES

Ao testar se uma correlação, r_1, é diferente de r_2 (a hipótese nula é de que $r_1 = r_2$; a hipótese alternativa é de que $r_1 \neq r_2$), para

$C1 = 0{,}5 \times \ln[(1 + r_1)/(1 - r_1)]$

$C2 = 0{,}5 \times \ln[(1 + r_2)/(1 - r_2)]$

Então:

$$N = [(Z_\alpha + Z_\beta) \div (C_1 - C_2)]^2 + 3.$$

APÊNDICE 6D

Tamanho de amostra para um estudo descritivo de uma variável contínua

TABELA 6D Tamanho de amostra para valores comuns de *A/DP**

	NÍVEL DE CONFIANÇA		
A/DP	**90%**	**95%**	**99%**
0,10	1.083	1.537	2.665
0,15	482	683	1.180
0,20	271	385	664
0,25	174	246	425
0,30	121	171	295
0,35	89	126	217
0,40	68	97	166
0,50	44	62	107
0,60	31	43	74
0,70	23	32	55
0,80	17	25	42
0,90	14	19	33
1,00	11	16	27

* *A/DP* é a amplitude padronizada do intervalo de confiança, computada como A (amplitude total desejada) dividido por DP (desvio-padrão da variável). Para estimar o tamanho total de amostra, localize a amplitude padronizada e o nível de confiança especificado.

FÓRMULA GERAL PARA OUTROS VALORES

Para outros valores de *A*, *DP*, e para um nível de confiança de (1 – α), o número total de sujeitos necessário (*N*) é

$$N = 4Z_{\alpha}^{2}DP^{2} \div A^{2}$$

(Veja Apêndice 6A para a definição de Z_{α}.)

APÊNDICE 6E

Tamanho de amostra para um estudo descritivo de uma variável dicotômica

TABELA 6E Tamanho de amostra para proporções

	NÚMERO SUPERIOR: NÍVEL DE CONFIANÇA DE 90% NÚMERO DO MEIO: NÍVEL DE CONFIANÇA DE 95% NÚMERO INFERIOR: NÍVEL DE CONFIANÇA DE 99%					
	AMPLITUDE TOTAL DO INTERVALO DE CONFIANÇA (A)						
PROPORÇÃO (*P*)* ESPERADA	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**
0,10	98	44	—	—	—	—	–
	138	61	—	—	—	—	–
	239	106	—	—	—	—	–
0,15	139	62	35	22	—	—	–
	196	87	49	31	—	—	–
	339	151	85	54	—	—	–
0,20	174	77	44	28	19	14	–
	246	109	61	39	27	20	–
	426	189	107	68	47	35	–
0,25	204	91	51	33	23	17	1
	288	128	72	46	32	24	1
	499	222	125	80	55	41	3
0,30	229	102	57	37	25	19	1
	323	143	81	52	36	26	2
	559	249	140	89	62	46	3
0,40	261	116	65	42	29	21	1
	369	164	92	59	41	30	2
	639	284	160	102	71	52	4
0,50	272	121	68	44	30	22	1
	384	171	96	61	43	31	2
	666	296	166	107	74	54	42

* Para estimar o tamanho de amostra, partindo da proporção esperada (*P*) para a variável de interesse, mova horizontalmente até a coluna da amplitude total desejada (*A*) para o intervalo de confiança. Os três valores representam o tamanho de amostra necessário para níveis de confiança de 90, 95 e 99%.

FÓRMULA GERAL PARA OUTROS VALORES

A fórmula geral para outros valores de *P*, *A*, e para um nível de confiança de (1 – α), onde *P* e *A* são definidos, é apresentada a seguir. Para:

z_α = desvio normalizado da distribuição normal para α bilateral, onde (1 – α) é o nível de confiança (p. ex., para um nível de confiança de 95%, α = 0,05 e Z_α = 1,96; para um nível de confiança de 90%, Z_α = 1,65; e para um nível de confiança de 99%, Z_α = 2,58).

Portanto, o número total de sujeitos necessário é:

$$N = 4Z_\alpha^2 \, P\,(1 - P) \div A^2$$

APÊNDICE 6F
Usos e abusos dos testes *t*

Os **testes *t* para duas amostras**, foco principal deste capítulo, são usados para comparar valores médios de um desfecho em dois grupos de sujeitos. Os dois grupos podem ser definidos por uma variável preditora (p. ex., medicamento ativo *versus* placebo em um ensaio clínico randomizado ou presença *versus* ausência de um fator de risco em um estudo de coorte), ou, então, por uma variável de desfecho em um estudo de caso-controle. Um teste *t* para duas amostras pode ser **não pareado**, se medidas obtidas em uma única ocasião estiverem sendo comparadas em dois grupos, ou **pareado**, se o que estiver sendo comparado entre os dois grupos for a mudança em um par de medidas realizadas em dois pontos do tempo (p. ex., antes e depois de uma intervenção). Um terceiro tipo de teste *t*, o **teste *t* pareado para uma única amostra**, compara a mudança média em um par de valores em um único grupo com uma mudança de zero ou com alguma outra mudança especificada.

A Tabela 6F mostra um uso inadequado do teste *t* pareado para uma única amostra em um ensaio clínico randomizado e cego que compara o efeito de um novo medicamento para dormir, em relação a um controle, na qualidade de vida. Em situações como essa, alguns pesquisadores produziram (e publicaram!) resultados utilizando dois testes *t* para uma amostra, separadamente para os grupos tratamento e controle.

Na tabela, os valores *P* assinalados com uma adaga (†) foram gerados pelos testes *t* pareados para uma única amostra. O primeiro valor *P* (0,05) revela uma mudança significativa na qualidade de vida nos sujeitos do grupo de tratamento; o segundo valor *P* (0,16) revela que não houve uma mudança significativa na qualidade de vida no grupo controle. Entretanto, essas análises não permitem fazer inferências sobre as diferenças na qualidade de vida entre os grupos e estaria incorreto concluir que existe um efeito significativo do tratamento com o novo medicamento.

Os valores *P* assinalados com um asterisco (*) representam os resultados apropriados produzidos por um teste *t* para duas amostras. Os dois primeiros valores *P* (0,87 e 0,64) correspondem aos testes *t* não

pareados que não mostram diferenças significativas entre os grupos nas medidas da qualidade de vida realizadas na linha de base e no final do estudo. O último valor *P* (0,17) foi produzido por um teste *t* pareado para duas amostras. Esse valor é mais próximo de 0,05 do que o valor *P* associado à medida da qualidade de vida do final do estudo (0,64) porque as médias das diferenças em cada par possuem desvios-padrão menores. Entretanto, a melhora na qualidade de vida no grupo tratado (1,3) não foi significativamente diferente do grupo placebo (0,9), e a conclusão correta é que o estudo não demonstra que o tratamento é efetivo.

TABELA 6F Maneiras corretas (e incorretas) de analisar dados pareados

MOMENTO DA REALIZAÇÃO DA AFERIÇÃO	QUALIDADE DE VIDA, MÉDIA ± DP		
	TRATAMENTO (N = 100)	CONTROLE (N = 100)	VALOR P
Linha de base	7,0 ± 4,5	7,1 ± 4,4	0,87*
Final do estudo	8,3 ± 4,7	8,0 ± 4,6	0,64*
Valor *P*	0,05†	0,16†	
Diferença	1,3 ± 2,1	0,9 ± 2,0	0,17*

* Comparando o tratamento com o controle.

† Comparando a linha de base com o final do estudo.

REFERÊNCIAS

1. Lehr R. Sixteen S-squared over D-squared: a relation for crude sample size estimates. *Stat Med* 1992;11:1099–1102.
2. Barthel FM, Babiker A, Royston P, Parmar MK. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Stat Med* 2006;25(15):2521–2542.
3. Ahnn S, Anderson SJ. Sample size determination in complex clinical trials comparing more than two groups for survival endpoints. *Stat Med* 1998;17(21):2525–2534.
4. Donner A. Sample size requirements for stratified cluster randomization designs [published erratum appears in *Stat Med* 1997;30(16):2927]. *Stat Med* 1992;11:743–750.
5. Kerry SM, Bland JM. Trials which randomize practices II: sample size. *Fam Pract* 1998;15:84–87.
6. Hemming K, Girling AJ, Sitch AJ, et al. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Med Res Methodol* 2011;11:102.
7. Jahn-Eimermacher A, Ingel K, Schneider A. Sample size in cluster-randomized trials

with time to event as the primary endpoint. *Stat Med* 2013;32(5):739–751.

8. Edwardes MD. Sample size requirements for case–control study designs. *BMC Med Res Methodol* 2001;1:11.
9. Drescher K, Timm J, Jöckel KH. The design of case–control studies: the effect of confounding on sample size requirements. *Stat Med* 1990;9:765–776.
10. Lui KJ. Sample size determination for case–control studies: the influence of the joint distribution of exposure and confounder. *Stat Med* 1990;9:1485–1493.
11. Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. *Stat Med* 2004;23(21):3263–3274.
12. Novikov I, Fund N, Freedman LS. A modified approach to estimating sample size for simple logistic regression with one continuous covariate. *Stat Med* 2010;29(1):97–107.
13. Vaeth M, Skovlund E. A simple approach to power and sample size calculations in logistic regression and Cox regression models. *Stat Med* 2004;23(11):1781–1792.
14. Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* 1998;19:589–601.
15. Murcray CE, Lewinger JP, Conti DV, et al. Sample size requirements to detect gene-environment interactions in genome-wide association studies. *Genet Epidemiol* 2011;35(3):201–210.
16. Wang S, Zhao H. Sample size needed to detect gene-gene interactions using linkage analysis. *Ann Hum Genet* 2007;71(Pt 6):828–842.
17. Witte JS. Rare genetic variants and treatment response: sample size and analysis issues. *Stat Med* 2012; 31(25):3041–3050.
18. Willan AR. Sample size determination for cost-effectiveness trials. *Pharmacoeconomics* 2011;29(11): 933–949.
19. Glick HA. Sample size and power for cost-effectiveness analysis (Part 2): the effect of maximum willingness to pay. *Pharmacoeconomics* 2011;29(4):287–296.
20. Glick HA. Sample size and power for cost-effectiveness analysis (Part 1). *Pharmacoeconomics* 2011;29(3):189–198.
21. Patel HI. Sample size for a dose-response study [published erratum appears in J Biopharm Stat 1994;4:127]. *J Biopharm Stat* 1992;2:l–8.
22. Day SJ, Graham DF. Sample size estimation for comparing two or more treatment groups in clinical trials. *Stat Med* 1991;10:33–43.
23. Guo JH, Chen HJ, Luh WM. Sample size planning with the cost constraint for testing superiority and equivalence of two independent groups. *Br J Math Stat Psychol* 2011;64(3):439–461.
24. Zhang P. A simple formula for sample size calculation in equivalence studies. *J Biopharm Stat* 2003;13(3):529–538.
25. Stucke K, Kieser M. A general approach for sample size calculation for the three-arm ‘gold standard’ non-inferiority design. *Stat Med* 2012;31(28):3579–3596.
26. Julious SA, Owen RJ. A comparison of methods for sample size estimation for

non-inferiority studies with binary outcomes. *Stat Methods Med Res* 2011;20(6):595–612.
27. Obuchowski NA. Sample size tables for receiver operating characteristic studies. *AJR Am J Roentgenol* 2000;175(3):603–608.
28. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763–770.
29. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85(3):257–268.
30. Jewell NP. *Statistics for epidemiology*. Boca Raton: Chapman and Hall, 2004, p. 68.

Delineamentos de pesquisa

Delineando estudos transversais e de coorte

Stephen B. Hulley, Steven R. Cummings e Thomas B. Newman

Os **estudos observacionais** têm dois objetivos principais: **descritivo**, examinar as distribuições das variáveis preditoras e de desfecho em uma população, e **analítico**, caracterizar as associações entre essas variáveis. Neste capítulo, apresentamos dois delineamentos observacionais básicos, definidos de acordo com a **sequência temporal** em que são feitas as aferições.

No **estudo transversal**, todas as medições são feitas em uma única ocasião ou durante um curto período de tempo. Sorteia-se uma amostra da população e examinam-se as distribuições das variáveis dentro dessa amostra, designando as variáveis preditora e de desfecho com base na plausibilidade biológica e em informações de outras fontes. Por exemplo, se o investigador estiver interessado em estudar a relação entre peso corporal e pressão arterial, ele poderia medir essas variáveis em cada participante em uma única visita à clínica onde é feita a pesquisa e avaliar se os sujeitos com maior peso corporal têm maior probabilidade de apresentarem hipertensão.

Em um **estudo de coorte**, as aferições ocorrem ao longo de um período de tempo em um grupo de participantes que foram identificados no início do estudo ("a coorte"). Assim, a característica que define os estudos de coorte é que **um grupo é reunido no início do estudo** e acompanhado **longitudinalmente**. Por exemplo, o investigador poderia medir o peso corporal e a pressão arterial em uma coorte de sujeitos de pesquisa em uma visita inicial à clínica e então acompanhá-los ao longo de cinco anos para determinar a relação entre o peso na linha de base e a incidência de hipertensão. Neste capítulo, discutimos delineamentos de **coorte prospectiva e retrospectiva**, assim como delineamentos de **coortes múltiplas**. Também discutimos as abordagens para **análise estatística** e a importância de otimizar a **retenção da coorte** durante o seguimento.

ESTUDOS TRANSVERSAIS

Em um estudo transversal, todas as medições são feitas em um único momento, sem período de seguimento (Figura 7.1). Os delineamentos transversais são úteis quando se quer descrever variáveis e seus padrões de distribuição. Por exemplo, no National Health and Nutrition Examination Survey (NHANES, Inquérito Nacional sobre Saúde e Nutrição), uma amostra que deveria representar toda a população dos Estados Unidos entre 1 e 74 anos foi entrevistada e examinada no início da década de 1970. Esse estudo transversal foi uma fonte importante de informações sobre a saúde e os hábitos da população norte-americana no ano em que foi realizado, fornecendo estimativas como a prevalência de tabagismo em vários grupos demográficos. Inquéritos do NHANES subsequentes têm sido feitos periodicamente, e todas as bases de dados do NHANES estão disponíveis para uso público (www.cdc.gov/nchs/nhanes.htm).

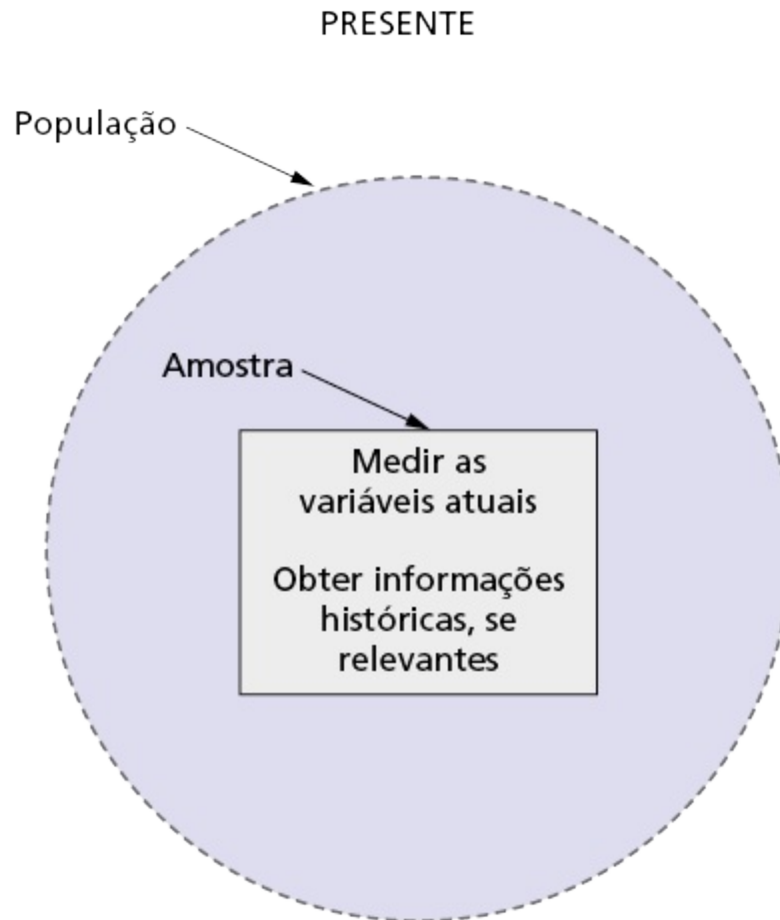



FIGURA 7.1 Em um estudo transversal, os passos são:

- Definir os critérios de seleção e recrutar uma amostra da população.
- Medir os valores atuais das variáveis preditoras e de desfecho, muitas vezes complementados por informações históricas.

Estudos transversais também podem examinar associações. Porém, nesse caso, as variáveis preditoras e de desfecho são definidas apenas com base nas hipóteses de causa-efeito do investigador, e não no delineamento do estudo. Essa definição é mais fácil para fatores constitucionais, como idade, raça e sexo, pois, como eles não podem ser alterados pelo efeito de outras variáveis, só podem ser considerados preditores. Para a maioria das variáveis, no entanto, a escolha se torna mais difícil. Por exemplo, o NHANES III encontrou associação transversal entre obesidade infantil e número de horas em que a criança assiste à televisão (1). Rotular a obesidade ou o fato de assistir à televisão como variável preditora e o outro como desfecho depende da hipótese causal do investigador.

Ao contrário dos estudos de coorte, que têm uma dimensão de tempo longitudinal e que podem ser usados para estimar a **incidência** (a proporção que *desenvolve* a doença ou condição clínica ao longo do tempo), os estudos transversais geralmente fornecem informações apenas sobre a **prevalência**, isto é, a proporção que *tem* a doença ou condição clínica em um determinado momento. A prevalência é útil para o médico que precisa estimar a probabilidade de o paciente que está sentado à sua frente ter uma certa doença; quanto maior a prevalência, maior a "probabilidade pré-teste" da doença (probabilidade antes de os resultados de diferentes testes diagnósticos estarem disponíveis; ver Capítulo 12). É por esse motivo que um número maior de pacientes com dor no joelho tem osteoartrose do que reumatismo palindrômico. A prevalência também é útil para o planejador da área da saúde que pretende saber quantas pessoas têm determinadas doenças, para que possa alocar recursos para cuidar dessas pessoas. Ao analisar estudos transversais, a prevalência do desfecho pode ser comparada entre os indivíduos com e sem uma exposição, fornecendo a **prevalência relativa** do desfecho, o equivalente para estudos transversais do risco relativo (ver Apêndice 8A para exemplos).

Às vezes, os estudos transversais descrevem a prevalência de já se ter feito algo ou de já se ter tido uma doença ou condição clínica. Nesse caso, é importante assegurar que o tempo de seguimento é o mesmo naqueles expostos e não expostos. Isso é ilustrado no Exemplo 7.1, no qual a prevalência de já se ter fumado foi pesquisada em um estudo transversal sobre crianças com diferentes níveis de exposição a filmes nos quais os atores fumam. Obviamente, as crianças que tinham visto mais filmes também eram mais velhas e, portanto, tiveram mais tempo para experimentarem o fumo; por esse motivo, foi muito importante ajustar para idade na hora de fazer as análises multivariadas (ver o Capítulo 9).

EXEMPLO 7.1 Estudo transversal

Para determinar se a exposição a filmes nos quais os atores fumam está associada a uma pessoa começar a fumar, Sargent e colaboradores (2):

1. Definiram os critérios de seleção e recrutaram a amostra. Fizeram

um inquérito com discagem de dígito aleatório de 6.522 crianças norte-americanas entre 10 e 14 anos de idade.

2. **Mediram as variáveis preditora e de desfecho.** Eles quantificaram o hábito de fumar em 532 filmes conhecidos e perguntaram para cada sujeito qual, de um subconjunto de 50 filmes escolhidos aleatoriamente, ele havia visto. Os sujeitos foram também indagados quanto a uma série de covariáveis como idade, raça, sexo, hábito tabagista e escolaridade dos pais, busca por sensações (p. ex., "gosto de fazer coisas perigosas") e autoestima (p. ex., "eu gostaria de ser outra pessoa"). A variável de desfecho foi se a criança já experimentou fumar um cigarro.

A prevalência de já ter fumado um cigarro variou de 2% no quartil inferior de exposição ao tabagismo nos filmes a 22% no quartil superior. Após ajustar para idade e outros confundidores, essas diferenças se mantiveram estatisticamente significativas; os autores estimaram que 38% da iniciação do hábito tabágico era atribuível à exposição a filmes nos quais os atores fumavam.

Pontos fortes e pontos fracos dos estudos transversais

Um ponto forte importante dos estudos transversais é que não é necessário esperar pela ocorrência do desfecho. Isso faz com que esses estudos sejam rápidos e de baixo custo e evita o problema das perdas no seguimento. Outra vantagem é que o estudo transversal pode ser incluído como primeiro passo em um estudo de coorte ou ensaio clínico com pouco ou nenhum custo adicional. Os resultados definem as características demográficas e clínicas do grupo de estudo na linha de base e podem às vezes revelar associações transversais interessantes.

Entretanto, como mencionado, geralmente é difícil estabelecer relações causais a partir de dados oriundos de um corte transversal no tempo. Os estudos transversais são também pouco práticos para estudar doenças raras, a não ser que a amostra tenha sido sorteada de uma população de pessoas doentes e não da população geral. Uma **série de casos** desse tipo funciona melhor para descrever características da doença do que para analisar diferenças entre esses pacientes e pessoas saudáveis. No entanto, às vezes é possível, por meio de comparações informais com a

experiência prévia, identificar fatores de risco muito fortes. Por exemplo, dos primeiros mil pacientes que tinham AIDS, 727 eram homens homossexuais ou bissexuais e 236 usavam drogas injetáveis (3). Não foi necessário um grupo-controle formal para concluir que esses grupos apresentavam maior risco. Além disso, é possível que haja associações de interesse clínico em uma amostra de pacientes com uma doença (p. ex., risco maior de sarcoma de Kaposi em pacientes com AIDS que eram homossexuais do que em usuários de drogas injetáveis).

Uma vez que os estudos transversais medem apenas a prevalência, e não a incidência, é preciso cautela ao buscar inferir sobre causalidade, prognóstico ou história natural de uma doença. Um fator que está associado a uma maior prevalência pode ser causa da doença, mas também pode estar associado a uma duração mais prolongada. Por exemplo, a prevalência de doença renal crônica depende não somente de sua incidência, mas também da sobrevida a partir do momento em que a doença surgiu. Devido à observação de que a obesidade está associada a uma maior sobrevida em pacientes em diálise (4), um estudo transversal sobre os preditores da doença renal crônica poderia superestimar a associação entre obesidade e doença renal crônica.

Séries (painéis) de inquéritos

Uma série de estudos transversais de uma única população observada em diversos momentos diferentes, por exemplo a cada 5 anos, pode ser usada para inferir sobre mudanças em padrões que variam com o tempo. Por exemplo, Zito e colaboradores (5), usando inquéritos transversais anuais, relataram que a prevalência do uso de medicações psicotrópicas em jovens com menos de 20 anos de idade aumentou mais de três vezes entre 1987 e 1996 em uma população do Medicaid na região do *mid-Atlantic,* nos Estados Unidos. **Séries de inquéritos transversais** têm um eixo temporal longitudinal, mas não são um estudo de coorte, pois uma nova amostra é sorteada a cada vez. Como resultado, não é possível avaliar mudanças em cada indivíduo, e os achados podem ser influenciados por pessoas entrando ou saindo da população (e, portanto, das amostras) devido a nascimentos, a óbitos e à migração.

ESTUDOS DE COORTE

Estudos de coorte prospectiva

Coorte era o termo usado na Roma antiga para definir um grupo de soldados que marchavam juntos. Na pesquisa clínica, coorte é um grupo de sujeitos especificados no início do estudo e seguidos no tempo. Em um **estudo de coorte prospectiva**, seleciona-se uma amostra de sujeitos (Figura 7.2) e, então, medem-se em cada sujeito características que poderão predizer desfechos subsequentes. A partir daí, esses sujeitos são seguidos no tempo por meio de medições periódicas dos desfechos de interesse (Exemplo 7.2).

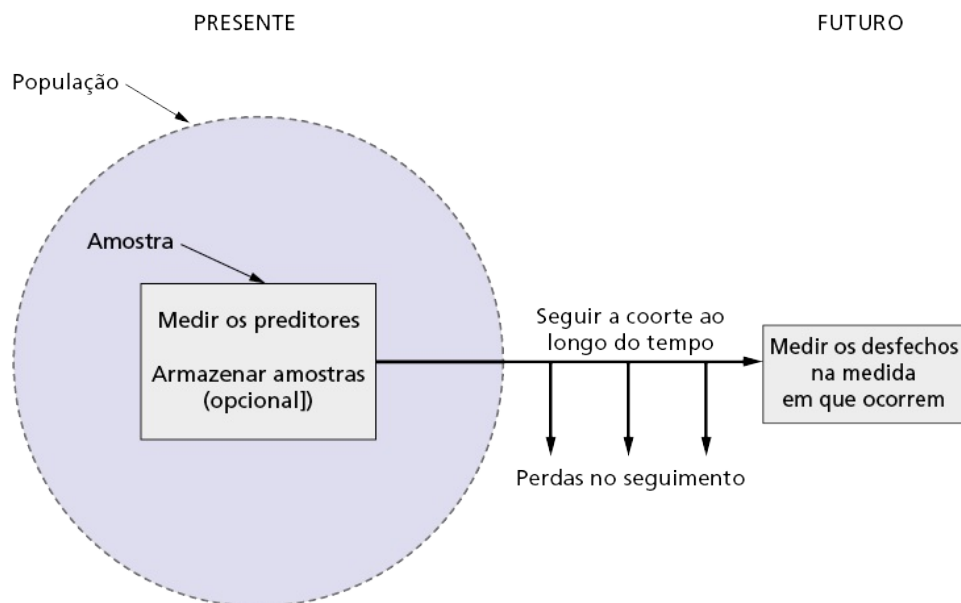



FIGURA 7.2 Em um estudo de coorte prospectiva, os passos são:

- Definir os critérios de seleção e recrutar uma amostra da população ("a coorte").
- Medir as variáveis preditoras e, se adequado, os níveis basais da variável de desfecho.
- Considerar a opção de armazenar amostras, imagens, etc. para análise posterior dos preditores.
- Seguir a coorte ao longo do tempo, minimizando perdas no seguimento.
- Medir as variáveis de desfechos durante o seguimento.

EXEMPLO 7.2 Estudo de coorte prospectiva

O estudo clássico do Nurses' Health Study avaliou a incidência e os fatores de risco para doenças comuns em mulheres. As etapas básicas do estudo foram:

1. **Definição dos critérios de seleção e montagem da coorte.** Em 1976, os investigadores obtiveram listas de enfermeiras com idades entre 25 e 42 anos nos 11 Estados mais populosos dos Estados Unidos e enviaram-lhes um convite para participar do estudo; as enfermeiras que aceitaram participar constituíram a coorte.
2. **Medição das variáveis preditoras, incluindo os potenciais confundidores.** Cada enfermeira recebeu um questionário sobre peso, nível de atividade física e outros potenciais fatores de risco. Um total de 121.700 enfermeiras retornaram os questionários preenchidos. Novos questionários foram enviados periodicamente, perguntando sobre outros fatores de risco e revisando o estado de alguns fatores de risco medidos anteriormente.
3. **Seguimento da coorte e medição dos desfechos.** Os questionários periódicos também incluíram questões sobre a ocorrência de uma série de desfechos, que foram validadas pelos investigadores.

Essa metodologia prospectiva permitiu aos investigadores realizar aferições na linha de base e coletar dados sobre desfechos subsequentes. O grande tamanho da coorte e o longo período de seguimento forneceram enorme poder estatístico para estudar os fatores de risco para cânceres e outras doenças.

Pontos fortes e pontos fracos dos estudos de coorte prospectiva

Uma vantagem importante do **delineamento de coorte** é que, ao contrário dos delineamentos transversais, ele permite calcular a **incidência** – número de casos novos de uma condição que surgem ao longo do tempo (Tabela 7.1). Como os níveis da variável preditora são aferidos antes da ocorrência do desfecho, esse delineamento permite também estabelecer a sequência temporal das variáveis, fortalecendo a inferência causal para uma associação. A abordagem prospectiva também impede que a aferição da variável preditora seja influenciada pelo desfecho ou pelo

conhecimento da ocorrência deste e permite ao investigador medir variáveis importantes de forma mais completa e acurada do que seria possível com uma abordagem retrospectiva. Isso pode ser importante, especialmente em estudos que avaliam certos tipos de preditores difíceis de serem lembrados corretamente, como hábitos alimentares. Além disso, quando doenças fatais são estudadas retrospectivamente, as variáveis preditoras nos indivíduos que morreram só podem ser reconstruídas a partir de fontes indiretas, como registros médicos ou relatos de amigos e familiares.

TABELA 7.1 Estatísticas para expressar a frequência de uma doença em estudos observacionais

TIPO DE ESTUDO	ESTATÍSTICA	DEFINIÇÃO
Transversal	Prevalência	Número de pessoas com uma doença ou condição em um dado momento ÷ Número de pessoas em risco
Coorte	Taxa de incidência	Número de pessoas que desenvolvem uma doença ou condição ÷ Número de pessoas em risco × período de tempo em risco

Todos os estudos de coorte apresentam uma mesma desvantagem, comum aos demais delineamentos observacionais, quando comparados aos ensaios clínicos: a inferência causal é difícil e a interpretação é frequentemente complicada pela influência de variáveis confundidoras (Capítulo 9). Um ponto especialmente vulnerável do delineamento prospectivo é que ele é uma forma cara e ineficiente para se estudar desfechos raros. Mesmo doenças que consideramos relativamente comuns, como câncer de mama, são, na verdade, tão raras de ano a ano, que um número muito grande de pessoas precisa ser seguido por longos períodos de tempo para que sejam observados desfechos em número suficiente para produzir resultados significativos. Os delineamentos de coorte são mais eficientes para desfechos dicotômicos comuns e imediatos e para desfechos contínuos.

Estudos de coorte retrospectiva

O delineamento do **estudo de coorte retrospectiva** (Figura 7.3) difere do delineamento prospectivo pelo fato de que a montagem da coorte, as

aferições da linha de base e o seguimento ocorreram no passado. Esse tipo de estudo só é possível se houver dados adequados disponíveis sobre as variáveis preditoras em uma coorte de sujeitos montada por outros motivos, como um banco eletrônico de dados clínicos ou administrativos (Exemplo 7.3).

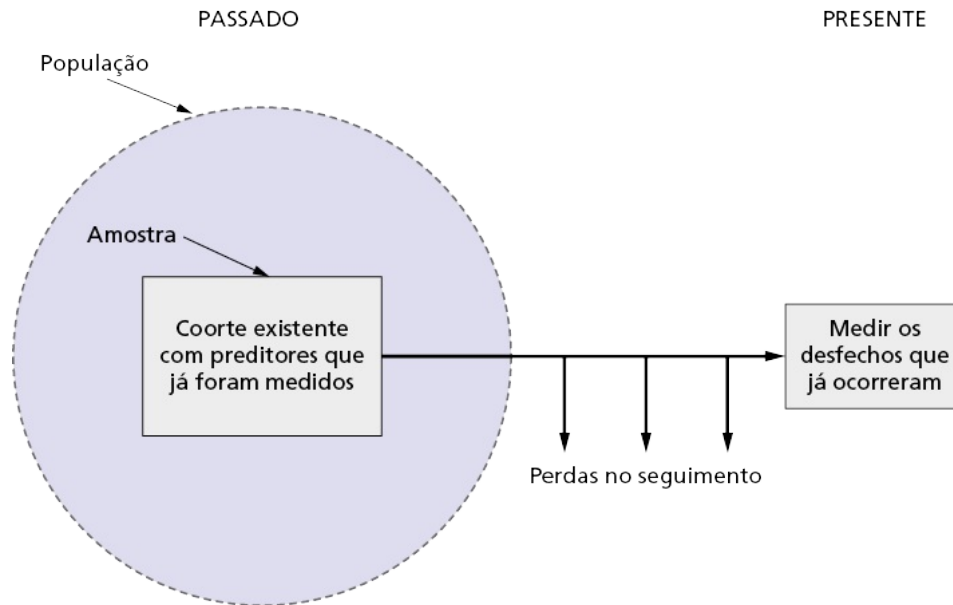



FIGURA 7.3 Em um estudo de coorte retrospectiva, a seleção da coorte e o seguimento ocorreram no passado; portanto, os passos são:

- Identificar uma coorte existente que já tenha algumas informações registradas sobre preditores.
- Avaliar a perda no seguimento que ocorreu.
- Medir as variáveis de desfecho que já ocorreram.

EXEMPLO 7.3 Estudo de coorte retrospectiva

Pearce et al. usaram dados do UK National Health Service Central Registry para descrever o risco de leucemia e de neoplasias cerebrais associado com tomografias computadorizadas de crânio na infância (7). As etapas do estudo foram:

1. **Identificação de uma coorte existente adequada.** A coorte consistiu em 178.604 crianças e jovens adultos com < 22 anos que realizaram TCs de crânio entre 1985 e 2002.
2. **Coleta de dados sobre as variáveis preditoras.** Os investigadores

revisaram os registros dos pacientes para averiguar sexo, idade, quantidade e tipos de procedimentos radiológicos e estimaram a dose de radiação a que os pacientes foram expostos.

3. **Coleta de dados sobre desfechos subsequentes.** Para evitar a inclusão de TCs relacionadas ao diagnóstico de câncer, os investigadores limitaram o estudo a casos de leucemia que ocorreram pelo menos dois anos após a primeira tomografia e tumores cerebrais que ocorreram pelo menos cinco anos após a primeira tomografia, identificando os casos diagnosticados até 2008.

As TCs de crânio realizadas na infância aumentaram significativamente o risco de leucemia e de neoplasias cerebrais, e o aumento foi relacionado à dose de radiação; doses cumulativas de 50 a 60 mGy triplicaram o risco de leucemia e de neoplasia cerebral. Entretanto, o aumento absoluto no risco de câncer foi baixo – um caso adicional de cada desfecho a cada 10.000 TCs de crânio realizadas. Os investigadores, embora reconhecendo que os benefícios da TCs de provavelmente superaram esses riscos, recomendaram fortemente que as doses de radiação emitidas pelas tomografias em crianças sejam mantidas no nível mais baixo possível e que procedimentos alternativos que evitem a emissão de radiação ionizante sejam considerados sempre que for apropriado.

Pontos fortes e pontos fracos

Os **estudos de coorte retrospectiva** têm muitos dos pontos fortes dos estudos de coorte prospectiva, tendo também a vantagem de serem muito mais baratos e de consumirem menos tempo. Os sujeitos já foram reunidos, as aferições basais já foram feitas e o período de seguimento já foi encerrado. As principais desvantagens são o controle limitado que o investigador tem sobre como delinear a estratégia de amostragem e o seguimento da população e sobre a natureza e a qualidade das aferições na linha de base. Os dados existentes podem ser incompletos, inacurados ou não terem sido medidos da forma ideal.

Estudos de coortes múltiplas e controles externos

Estudos de coortes múltiplas começam com duas ou mais amostras separadas de sujeitos: em geral, um grupo **exposto** a um potencial fator de risco e um ou mais grupos não expostos ou com níveis mais baixos de exposição (Figura 7.4). Uma vez definidas coortes apropriadas para o estudo que tenham níveis diferentes de exposição ao preditor de interesse, procede-se à medição de outras variáveis preditoras, ao seguimento da coorte e à aferição dos desfechos, como em qualquer outro tipo de estudo de coorte.

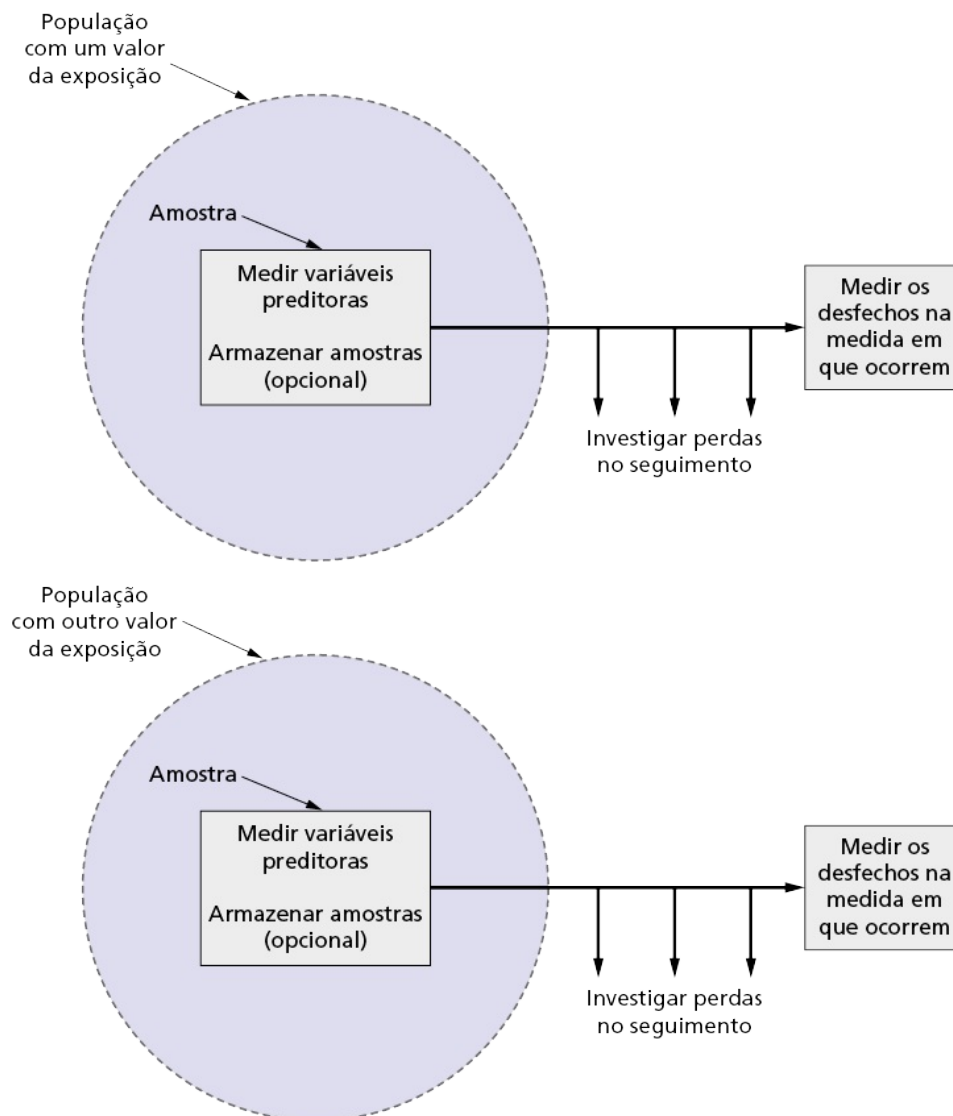



FIGURA 7.4 Em um estudo de dupla coorte (que pode ser conduzido de forma prospectiva ou retrospectiva), os passos são:

- Selecionar duas ou mais coortes com níveis diferentes da exposição

(preditor principal).

- Medir outros preditores.
- Medir variáveis de desfecho no seguimento.

EXEMPLO 7.4 Delineamento de coortes múltiplas

Para averiguar se um grau elevado de icterícia neonatal ou de desidratação tem efeitos deletérios importantes sobre o desenvolvimento neurológico, investigadores da UCSF e do Kaiser Permanente do norte da Califórnia realizaram um estudo de tripla-coorte (8, 9). As etapas básicas do estudo foram:

1. **Identificação de coortes com níveis diferentes de exposição.** Os investigadores usaram bases de dados eletrônicas para identificar recém-nascidos a termo e próximos do termo que
   1. tinham nível sérico total de bilirrubinas ≥ 25 mg/dL, ou
   2. foram readmitidos por desidratação com um sódio sérico ≥ 150 mEq/L ou perda de peso ≥ 12% desde o nascimento, ou
   3. foram selecionados aleatoriamente da coorte de nascimentos.
2. **Coleta de dados sobre os desfechos.** Os investigadores usaram bases de dados eletrônicas para buscar os diagnósticos de doenças neurológicas e realizaram um exame neurocomportamental completo aos cinco anos de idade naqueles que consentiram em participar (cegados para a qual das três coortes o participante pertencia).

A hiperbilirrubinemia e a desidratação não estiveram associadas a efeitos deletérios na saúde dessas crianças.

Embora o **delineamento de dupla-coorte** use duas amostras diferentes de sujeitos, não deve ser confundido com o estudo de caso-controle (Capítulo 8). As amostras em um estudo de dupla-coorte são escolhidas com base nos níveis de uma determinada variável preditora. Por outro lado, em um estudo de caso-controle, as amostras são escolhidas com base na presença ou na ausência do desfecho.

Uma variante do delineamento de coortes múltiplas é comparar os desfechos de membros de uma coorte com dados de um **censo ou registro**

de uma população diferente. Um exemplo é um estudo clássico de Wagoner e colaboradores (6). Para determinar se os trabalhadores de minas de urânio apresentam uma incidência mais elevada de câncer de pulmão, os autores compararam a incidência de câncer respiratório em 3.415 trabalhadores de minas de urânio com a de homens brancos nos mesmos Estados. A maior incidência de câncer de pulmão observada nos trabalhadores contribuiu para estabelecer que a exposição ocupacional à radiação ionizante é uma causa importante de câncer de pulmão.

Pontos fortes e pontos fracos dos estudos de coortes múltiplas

O delineamento de coortes múltiplas pode ser a única forma factível de estudar exposições raras e exposições a potenciais fatores de risco ocupacionais ou ambientais. O uso de dados de censos ou registros para constituir um grupo-controle externo tem as vantagens adicionais de dar uma base populacional e de ser mais econômico. Afora isso, os pontos fortes desse delineamento são semelhantes aos dos demais estudos de coorte.

O problema do **confundimento** é acentuado em estudos de coortes múltiplas, pois as coortes são montadas a partir de populações diversas que podem diferir de formas importantes (por outros fatores que não a exposição à variável preditora) a ponto de influenciar os desfechos. Embora algumas dessas diferenças, como idade e raça, possam ser pareadas ou usadas para ajustar os achados estatisticamente, outras características importantes poderão não estar disponíveis e, assim, gerar problemas na hora de interpretar as associações observadas.

ABORDAGEM ESTATÍSTICA EM ESTUDOS DE COORTE

Riscos, **chances** (*odds*) e **taxas** são estimativas da frequência de um desfecho dicotômico em sujeitos que foram acompanhados por um período de tempo. Essas três medidas estão fortemente relacionadas entre si, compartilhando o mesmo numerador – o número de sujeitos que desenvolveram um desfecho dicotômico. Nessas três medidas está implícito o conceito de estar *em risco*, que implica que, no início do estudo, o sujeito ainda não havia desenvolvido o desfecho de interesse. Em um estudo prospectivo sobre os fatores de risco para o

desenvolvimento do diabetes, não se pode afirmar que uma mulher que já tinha diabetes na linha de base está em risco, pois ela já desenvolveu o desfecho de interesse. Por outro lado, há condições episódicas, como descompensação de insuficiência cardíaca requerendo internação hospitalar, nas quais o desfecho de interesse pode ser a incidência de um novo episódio, mesmo se ele ocorrer em alguém que já o teve anteriormente.

Considere um estudo que acompanhou 1.000 pessoas ao longo de dois anos para ver quem desenvolvia câncer de pulmão, no qual ocorreram oito casos novos a cada ano. O risco, as chances e a taxa são mostrados na Tabela 7.2.

TABELA 7.2 Cálculos de risco, chances e taxa em um estudo que acompanhou 1.000 pessoas ao longo de dois anos, tendo ocorrido oito casos novos de câncer de pulmão a cada ano

ESTATÍSTICA	FÓRMULA	EXEMPLO
Risco	*N de pessoas que desenvolveram o desfecho* ÷ *N de pessoas em risco*	16 ÷ 1.000 = 0,016
Chances (odds)	*N de pessoas que desenvolveram o desfecho* ÷ *N de pessoas que não desenvolveram o desfecho*	16 ÷ 984 = 0,0163
Taxa*	*N de pessoas que desenvolveram o desfecho* ÷ *Pessoas-tempo em risco*	16 casos ÷ 1.992 pessoas-ano = 0,008 casos/pessoas-ano

* O denominador para a taxa é o número de pessoas em risco no primeiro ano (1.000) somado ao número em risco no segundo ano (992).

Dessas três medidas, o risco é a mais fácil de compreender, pois é a mais familiar para a maioria das pessoas – o risco de desenvolver câncer de pulmão em dois anos foi de 16 em 1.000. As chances são a medida mais difícil de compreender intuitivamente – as chances de desenvolver câncer de pulmão foram de 16 em 984; felizmente, para desfechos raros (como nesse caso), as chances são quantitativamente semelhantes ao risco e não apresentam vantagem sobre ele. Em estudos que comparam dois grupos, *quando o desfecho é raro*, a **razão de chances** é também

semelhante à **razão de riscos** (risco relativo). Esse fato tem grande importância em duas situações: a razão de chances é a base para os cálculos da regressão logística e é usada para obter uma estimativa aproximada do risco relativo em estudos de caso-controle (Apêndice 8B). As taxas, que levam em consideração o acúmulo de eventos ao longo do tempo, são expressas como o número de eventos dividido pelo número de pessoas-tempo em risco de desenvolvê-lo, ou seja, pelo tempo total de seguimento para cada um dos sujeitos do estudo, desde que o indivíduo esteja vivo, permaneça no estudo e ainda não tenha desenvolvido o desfecho.

Os estudos de coorte podem ser comprometidos pela **perda importante de sujeitos no seguimento**, por tempos de seguimento desiguais entre os grupos ou pela ocorrência de óbitos ou outros eventos que impedem a aferição do desfecho. Nesses casos, é útil comparar as **taxas de incidência** entre os grupos – o número de desfechos dividido pelo número de pessoas-tempo em risco. Cada sujeito no estudo contribui com meses ou anos de pessoa-tempo, desde sua entrada na coorte até desenvolver o desfecho de interesse ou ser "**censurado**" devido à perda no seguimento ou ao óbito. A taxa de incidência em qualquer grupo do estudo é o número de desfechos no grupo dividido pelo somatório das pessoas-tempo em risco naquele grupo. Assim como ocorre na **razão de riscos** (também conhecida como risco relativo), pode-se estimar a **razão de taxas** como o quociente entre as taxas em pessoas que têm e em pessoas que não têm um determinado fator de risco. O modelo de azares proporcionais de Cox fornece um método para a análise multivariada de dados desse tipo (às vezes chamados dados tipo "tempo até o evento"); esse método permite estimar a **razão de azares**, que é similar à razão de taxas e ganhou popularidade como medida de associação nas **análises de regressão de Cox**.

Outros tópicos relacionados a estudos de coorte

A característica básica de um estudo de coorte é a definição de um grupo de sujeitos no início do período de seguimento. Os sujeitos devem ser apropriados à questão de pesquisa e estar disponíveis para o seguimento. Devem, também, ser suficientemente semelhantes à população para a qual os resultados serão generalizados. Além disso, o número de sujeitos

recrutados deve fornecer poder estatístico adequado.

A qualidade do estudo dependerá da precisão e da acurácia das aferições das variáveis preditoras e de desfecho. A capacidade de inferir sobre causa e efeito depende também da identificação e aferição de todos os **potenciais confundidores** (Capítulo 9), e a capacidade de generalizar para subgrupos da população depende da aferição de todas as possíveis **fontes de modificação de efeito**. As variáveis preditoras podem mudar durante o estudo, e a decisão sobre repetir ou não as medições e a sua frequência depende do custo, da probabilidade de mudanças na variável e da importância que a observação dessas mudanças tem para a questão de pesquisa. Os desfechos devem ser avaliados com critérios padronizados e, quando sua avaliação puder ser influenciada pelo conhecimento acerca dos fatores de risco, recomenda-se que as pessoas que fazem essas avaliações estejam cegadas para as variáveis preditoras.

A **capacidade de seguir a coorte inteira** é importante, e os estudos prospectivos devem adotar uma série de medidas para alcançar essa meta (Tabela 7.3). Os sujeitos que planejam se mudar para lugares de difícil acesso durante o estudo ou cujo seguimento será dificultado por outros motivos devem ser excluídos desde o início. No momento do ingresso no estudo, o investigador deve coletar informações que permitam a localização em caso de mudança ou de falecimento, como endereço, número de telefone e e-mail do sujeito, de seu médico e de um ou dois amigos próximos ou familiares que não morem com o sujeito. Números de telefone celular e endereços de e-mail são particularmente úteis, pois, em geral, não mudam quando os sujeitos, amigos ou familiares mudam de endereço ou trocam de emprego. Se possível, obter algum número de identidade irá ajudar na determinação do estado vital de sujeitos perdidos no seguimento e na obtenção de dados de alta hospitalar. O contato periódico com os sujeitos uma ou duas vezes por ano ajuda a manter seu vínculo no estudo, a aumentar a acurácia dos desfechos de interesse e a precisar melhor o momento da sua ocorrência. Localizar os sujeitos para avaliações de seguimento às vezes pode demandar esforços persistentes e repetidos por correio, e-mail, telefone, ou até mesmo visita domiciliar.

TABELA 7.3 Estratégias para minimizar perdas no seguimento

Durante o arrolamento

1. Excluir sujeitos com alta probabilidade de perda
   a. Aqueles que planejam se mudar
   b. Aqueles que podem não estar dispostos a retornar
   c. Aqueles com problema de saúde ou doença fatal não relacionada à questão de pesquisa
2. Obter informações que permitam futura localização
   a. Endereço, número de telefone (números de telefone celular são particularmente úteis) e e-mail do sujeito
   b. Número de identidade (p. ex., *Social Security/Medicare**)
   c. Nome, endereço, número de telefone e e-mail de amigos próximos ou parentes que não moram com o sujeito
   d. Nome, e-mail, endereço e número de telefone do médico que acompanha a pessoa

Durante o seguimento†

1. Contato periódico com os sujeitos para coletar informações, fornecer resultados e oferecer apoio:
   a. Por telefone: pode ser necessário ligar durante finais de semana e à noite
   b. Por correio: envios repetidos de mensagens por e-mail ou de cartões selados autoendereçados por correio normal
   c. Outros: jornais, brindes com o logo do estudo
2. Para os que não podem ser localizados por telefone ou correio:
   a. Contatar amigos, parentes ou médicos
   b. Solicitar o novo endereço por meio de serviços postais
   c. Buscar o endereço por outras fontes, como listas telefônicas, Internet e, em último caso, serviços de crédito
   d. Para sujeitos com *Medicare*, coletar dados sobre altas hospitalares da *Social Security Administration*
   e. Determinar o estado vital a partir do State Health Department ou do National Death Index***

Em todos os momentos

1. Tratar os sujeitos do estudo com atenção, carinho e respeito, ajudando-os a compreender a questão de pesquisa de modo que queiram atuar como parceiros para o sucesso do estudo.

† Isso pressupõe que os participantes do estudo tenham fornecido o consentimento informado para a coleta de informações de rastreamento e para os contatos de seguimento.
* N. de R. T. No Brasil, idealmente por meio do CPF.
** N. de R. T. No Brasil, as informações sobre altas hospitalares do SUS podem ser obtidas por meio do Sistema de Informações Hospitalares (SIH/SUS).
*** N. de R. T. No Brasil, pela Declaração de óbito junto às secretarias municipais ou estaduais de saúde.

RESUMO

1. Em um **estudo transversal**, as variáveis são todas medidas em uma única vez, sem distinção estrutural entre as preditoras e as de desfecho. Os estudos transversais produzem **evidências mais fracas a respeito de causalidade** do que os estudos de coorte, pois não demonstram que a variável preditora precede o desfecho.
2. Os estudos transversais são valiosos para fornecer informações descritivas sobre **prevalência** e têm a vantagem de **poupar tempo**, **recursos financeiros** e evitar o **problema dos abandonos**, típicos de um estudo de seguimento. Muitas vezes são úteis como o primeiro passo de um estudo de coorte ou experimental. É possível também juntar estudos transversais na forma de **séries de inquéritos** com amostragem independente, para revelar mudanças populacionais que

ocorrem ao longo do tempo.

3. Os estudos transversais requerem um tamanho de amostra muito grande para estudar doenças e variáveis que são raras na população geral, mas podem ser usados para estudar uma **série de casos** de uma doença rara.
4. Em **estudos de coorte**, um grupo de sujeitos identificado no início do estudo é acompanhado ao longo do tempo para descrever a incidência ou a história natural de uma condição e para identificar preditores (fatores de risco) para vários desfechos. A capacidade de medir a variável preditora antes da ocorrência do desfecho permite estabelecer a sequência de eventos e controlar os vieses em sua aferição.
5. Estudos de **coorte prospectiva** começam no início do seguimento e podem demandar números elevados de sujeitos seguidos durante longos períodos de tempo. Essa desvantagem pode ser superada identificando-se uma **coorte retrospectiva** na qual as aferições das variáveis preditoras já foram feitas.
6. O delineamento de **coortes múltiplas**, que compara a incidência dos desfechos em coortes que diferem quanto ao nível de uma variável preditora (“a **exposição**”), é útil para estudar os efeitos de exposições raras e ocupacionais.
7. **Riscos**, **chances** (*odds*) e **taxas** são três formas de estimar a frequência de um desfecho dicotômico durante o seguimento; as **taxas de incidência**, que levam em conta o número de pessoas-tempo entre os participantes que permanecem vivos e livres de eventos no estudo, são a base para abordagens modernas de cálculo das **razões de azares multivariadas**, por meio dos modelos de azares proporcionais de Cox.
8. As inferências sobre **causa e efeito** são fortalecidas medindo e ajustando para todas as possíveis variáveis confundidoras. O viés na avaliação dos desfechos pode ser prevenido **padronizando** as aferições e **cegando** as pessoas responsáveis pela avaliação do desfecho quanto aos valores das variáveis preditoras.
9. Os pontos fortes de um delineamento de coorte podem ser comprometidos pelo **seguimento** incompleto dos sujeitos. As perdas podem ser minimizadas **excluindo sujeitos** que se identifica desde o início que poderão não estar disponíveis para o seguimento, coletando **informações na linha de base** que facilitem a localização e

mantendo-se em contato com todos os sujeitos regularmente.

REFERÊNCIAS

1. Andersen RE, Crespo CJ, Bartlett SJ, et al. Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. *JAMA* 1998;279(12):938–942.
2. Sargent JD, Beach ML, Adachi-Mejia AM, et al. Exposure to movie smoking: its relation to smoking initiation among US adolescents. *Pediatrics* 2005;116(5):1183–1191.
3. Jaffe HW, Bregman DJ, Selik RM. Acquired immune deficiency syndrome in the United States: the first 1,000 cases. *J Infect Dis* 1983;148(2):339–345.
4. Kalantar-Zadeh K, Abbott KC, Salahudeen AK, et al. Survival advantages of obesity in dialysis patients. *Am J Clin Nutr* 2005; 81: 543–554.
5. Zito JM, Safer DJ, DosReis S, et al. Psychotropic practice patterns for youth: a 10-year perspective. *Arch Pediatr Adolesc Med* 2003;157(1):17–25.
6. Huang Z, Hankinson SE, Colditz GA, et al. Dual effect of weight and weight gain on breast cancer risk. *JAMA* 1997;278:1407–1411.
7. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukemia and brain tumors: a retrospective cohort study. *Lancet* 2012;380:499–505.
8. Newman TB, Liljestrand P, Jeremy RJ, et al. Outcomes of newborns with total serum bilirubin levels of 25 mg/dL or more. *N Engl J Med* 2006;354:1889–1900.
9. Escobar GJ, Liljestrand P, Hudes ES, et al. Five-year neurodevelopmental outcome of neonatal dehydration. *J Pediatr* 2007;151(2):127–133, 133 e1.
10. Wagoner JK, Archer VE, Lundin FE, et al. Radiation as the cause of lung cancer among uranium miners. *N Engl J Med* 1965;273:181–187.

Delineando estudos de caso-controle

Thomas B. Newman, Warren S. Browner, Steven R. Cummings e Stephen B. Hulley

No Capítulo 7, introduzimos os estudos de coorte, em que a sequência das medições é a mesma da cronologia de causa e efeito: primeiro, mede-se a variável preditora e, depois (após um intervalo de seguimento), o desfecho. O **estudo de caso-controle** inverte essa sequência temporal. Ele começa com a seleção de uma amostra de pacientes com o desfecho (casos) e outra amostra sem o desfecho (controles); então, comparam-se os níveis das variáveis preditoras nas duas amostras para determinar quais estão associadas ao desfecho. Por exemplo, um estudo de caso-controle poderia reunir um conjunto de casos de melanoma ocular e uma amostra de controles hígidos. Então, buscar-se-iam dados em cada grupo sobre exposição prévia à soldagem a arco elétrico, para estimar como ela afeta o risco de desenvolver melanoma ocular. O delineamento de caso-controle tem **custo relativamente baixo** e é muito **eficiente** para estudar doenças raras.

Este capítulo também apresenta diversas variações do delineamento de caso-controle básico descrito. O delineamento de **caso-controle** *aninhado* compara os casos incidentes aninhados em um estudo de coorte com controles sorteados aleatoriamente do resto da coorte; esse delineamento controla o viés de amostragem e de aferição e economiza recursos financeiros se as variáveis preditoras dependerem de aferições caras que podem ser feitas em material biológico armazenado ou em imagens coletadas no início do estudo de coorte. O delineamento de **caso-controle com amostragem por** *incidência-densidade* permite analisar relações de risco, levando em consideração as mudanças que ocorrem ao longo do

tempo nos níveis dos fatores de risco e as perdas no seguimento. Por fim, o delineamento de **caso-coorte aninhado** permite sortear uma amostra aleatória de toda a coorte para servir como controle para diferentes conjuntos de casos. O capítulo termina com sugestões sobre como escolher entre os diferentes delineamentos observacionais discutidos nos Capítulos 7 e 8.

ESTUDOS DE CASO-CONTROLE

Como a maioria das doenças são relativamente incomuns, os estudos de coorte ou transversais de amostras da população geral apresentam custo elevado, podendo requerer milhares de sujeitos para identificar fatores de risco para uma doença rara como o câncer gástrico. Como discutido no Capítulo 7, uma **série de casos** de pacientes com a doença, por sua vez, poderia identificar um fator de risco óbvio (como, p. ex., para a AIDS, o uso de drogas injetáveis), a partir de conhecimentos prévios sobre a prevalência do fator de risco na população geral. Para a maioria dos fatores de risco, no entanto, é necessário montar um grupo-controle, de referência, para que a prevalência do fator de risco nos sujeitos com a doença (casos) possa ser comparada com a prevalência em sujeitos sem a doença (controles).

Os estudos de caso-controle são **retrospectivos** (Figura 8.1). O estudo identifica um grupo de sujeitos com a doença e outro sem, e então olha para o passado para identificar diferenças nas variáveis preditoras que possam explicar por que os casos desenvolveram a doença e os controles não (Exemplo 8.1).

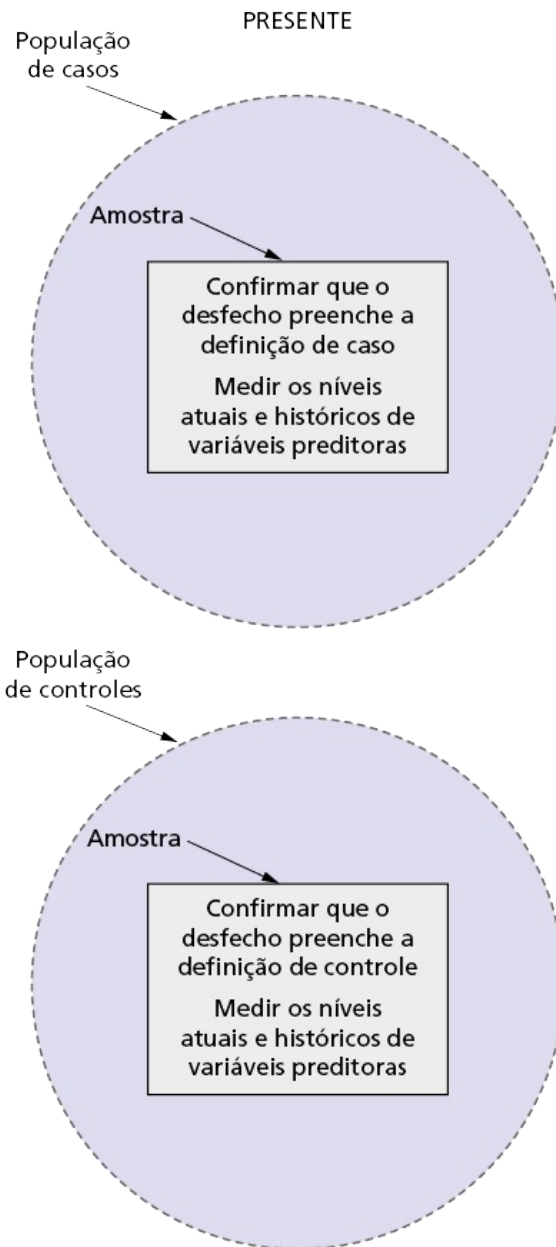



FIGURA 8.1 Em um estudo de caso-controle, os passos são:

- Definir os critérios de seleção e recrutar uma amostra de uma população de casos e uma segunda amostra de uma população de controles.
- Medir os valores atuais de variáveis relevantes, muitas vezes complementando com informações históricas.

EXEMPLO 8.1 Estudo de caso-controle

Como a vitamina K intramuscular (IM) é administrada rotineiramente

em recém-nascidos nos Estados Unidos, dois estudos que mostraram um aumento de duas vezes no risco de câncer infantil em crianças que receberam vitamina K por via intramuscular causaram bastante polêmica (4, 5). Para investigar essa associação mais detalhadamente, investigadores alemães (6):

1. **Selecionaram a amostra de casos.** 107 crianças com leucemia nos Registros de Câncer Infantil da Alemanha.
2. **Selecionaram a amostra de controles.** 107 crianças pareadas por sexo e data de nascimento e selecionadas aleatoriamente entre as crianças que moravam na mesma cidade que os casos no momento do diagnóstico (com base nos registros residenciais fornecidos pela prefeitura).
3. **Mediram a variável preditora.** Revisaram registros médicos para determinar quais casos e controles haviam recebido vitamina K IM logo após o nascimento.

Os autores descobriram que 69 dos 107 casos (64%) e 63 dos 107 controles (59%) haviam sido expostos à vitamina K IM, com uma razão de chances (*odds ratio*) de 1,3 (intervalo de confiança [IC] de 95%, 0,7 a 2,3). (Veja Apêndice 8A para o cálculo.) Portanto, esse estudo não confirmou a associação entre a aplicação de vitamina K IM logo após o nascimento e a subsequente leucemia infantil. A estimativa-ponto e o limite superior do IC de 95% deixaram em aberto a possibilidade de um aumento clinicamente importante da leucemia na população da qual foram selecionadas as amostras, porém vários outros estudos e uma análise usando um grupo-controle adicional no estudo descrito nesse exemplo também não confirmaram a associação (7, 8).

Os estudos de caso-controle surgiram como estudos epidemiológicos que buscavam identificar fatores de risco para doenças. Por esse motivo, e também para facilitar a discussão, consideramos como “casos” aqueles indivíduos que têm a doença. No entanto, o delineamento de caso-controle também pode ser usado para examinar outros desfechos incomuns, como invalidez em indivíduos já doentes. Além disso, quando desfechos indesejados são a regra e não a exceção, os casos em um estudo

de caso-controle podem ser aqueles raros pacientes com um bom desfecho, como a recuperação de uma doença normalmente fatal.

Os estudos de caso-controle são o “tinto da casa” da lista de vinhos de delineamentos de pesquisa: são mais modestos e um pouco mais arriscados do que as outras seleções, mas muito menos caros e, às vezes, surpreendentemente bons. O delineamento de um estudo de caso-controle é desafiador devido às várias oportunidades para a ocorrência de vieses, porém, há muitos exemplos de estudos de caso-controle bem-delineados que produziram resultados importantes. Eles incluem as associações entre o uso materno de dietilestilbestrol (DES) e câncer vaginal na prole feminina (um estudo clássico que forneceu uma conclusão definitiva com base em apenas sete casos) (1) e a posição de decúbito ventral e sua relação com a síndrome da morte súbita do lactente (2), um resultado simples que salvou a vida de milhares de crianças (3).

Os estudos de caso-controle não podem produzir estimativas sobre a incidência ou prevalência de uma doença, pois a proporção de sujeitos com a doença no estudo é determinada pelo número de casos e controles que o investigador decide amostrar, e não pelas suas proporções na população. O que os estudos de caso-controle podem fornecer são informações descritivas sobre as características dos casos e, o que é mais importante, uma estimativa da magnitude da associação entre cada variável preditora e o desfecho. Essas estimativas são expressas na forma de razões de chances (*odds ratios*, razões de *odds*), que se aproximam do risco relativo se o risco da doença em pessoas expostas e não expostas for relativamente baixo (em torno de 10% ou menos) (Apêndice 8B).

Pontos fortes dos estudos de caso-controle

Eficiência para desfechos raros

Uma das principais vantagens dos estudos de caso-controle é o grande número de informações que podem ser fornecidas rapidamente a partir de um número relativamente pequeno de sujeitos. Considere um estudo sobre o efeito da circuncisão no carcinoma subsequente do pênis. Esse câncer é muito raro em homens circuncidados, mas também é raro em homens não circuncidados. A sua incidência cumulativa ao longo da vida do indivíduo é de aproximadamente 0,16% (9). Para realizar um estudo de coorte com

uma probabilidade razoável (80%) de detectar um fator de risco muito forte (p. ex., um risco relativo de 50), seria necessário incluir mais de 6 mil homens, pressupondo proporções grosseiramente iguais de circuncidados e não circuncidados. Um ensaio clínico randomizado sobre circuncisão no nascimento exigiria o mesmo tamanho de amostra, mas os casos ocorreriam em média 67 anos após o ingresso no estudo. Seriam necessárias três gerações de investigadores para acompanhar os sujeitos.

Agora considere um estudo de caso-controle sobre a mesma questão. Para a mesma probabilidade de detecção do mesmo risco relativo, seriam necessários apenas 16 casos e 16 controles (e não muito tempo despendido pelo investigador). Para doenças raras ou com longos períodos de latência entre a exposição e a doença, os estudos de caso-controle são muito mais eficientes do que os outros delineamentos. Muitas vezes, inclusive, são a única opção factível.

Utilidade na geração de hipóteses

A abordagem retrospectiva dos estudos de caso-controle e sua capacidade de examinar um grande número de variáveis preditoras os tornam úteis para a geração de hipóteses sobre as causas de um novo surto de doença. Por exemplo, um estudo de caso-controle sobre uma epidemia de insuficiência renal aguda em crianças do Haiti encontrou uma razão de chances de 53 para a ingestão de xarope de paracetamol fabricado na região. Investigações mais detalhadas revelaram que a insuficiência renal foi causada pela intoxicação por dietilenoglicol, que contaminou o xarope de paracetamol (10), um problema infelizmente recorrente (11).

Pontos fracos dos estudos de caso-controle

Apesar das vantagens dos estudos de caso-controle, eles apresentam também algumas limitações importantes. Em primeiro lugar, é possível estudar apenas um único desfecho (presença ou ausência da doença, que foi também o critério para selecionar as duas amostras), ao passo que os estudos de coorte e transversais (e também os ensaios clínicos) permitem estudar diversas variáveis de desfecho. Em segundo lugar, como já mencionado, as informações que os estudos de caso-controle podem fornecer são limitadas: não há como estimar diretamente a incidência ou prevalência da doença, nem o risco atribuível ou o excesso de risco, a não

ser que haja dados sobre a população e sobre o momento no tempo em que os casos surgiram. A maior limitação dos estudos de caso-controle, porém, é a grande **suscetibilidade a vieses**. Esses vieses provêm principalmente de duas fontes: **amostragem separada** dos casos e dos controles e **aferição retrospectiva** das variáveis preditoras. Essas duas fontes de vieses e as estratégias para lidar com elas são abordadas nas próximas duas seções.

Viés de amostragem e como controlá-lo

A amostragem em um estudo de caso-controle inicia com os casos. Idealmente, a amostra dos casos deve incluir todos os indivíduos que desenvolveram a doença em estudo ou um conjunto selecionado aleatoriamente dessa amostra maior. No entanto, surge um problema imediato. Como saber quem desenvolveu a doença e quem não a desenvolveu? Em estudos transversais e de coorte, a doença é buscada sistematicamente em todos os participantes do estudo, porém, nos estudos de caso-controle, os casos devem ser amostrados de pacientes nos quais a doença já foi diagnosticada e que estão disponíveis para estudo. Essa amostra pode não ser representativa de todos os pacientes que desenvolveram a doença, uma vez que aqueles que não foram diagnosticados, foram mal diagnosticados, não estão disponíveis ou já faleceram provavelmente não serão incluídos (Figura 8.2).

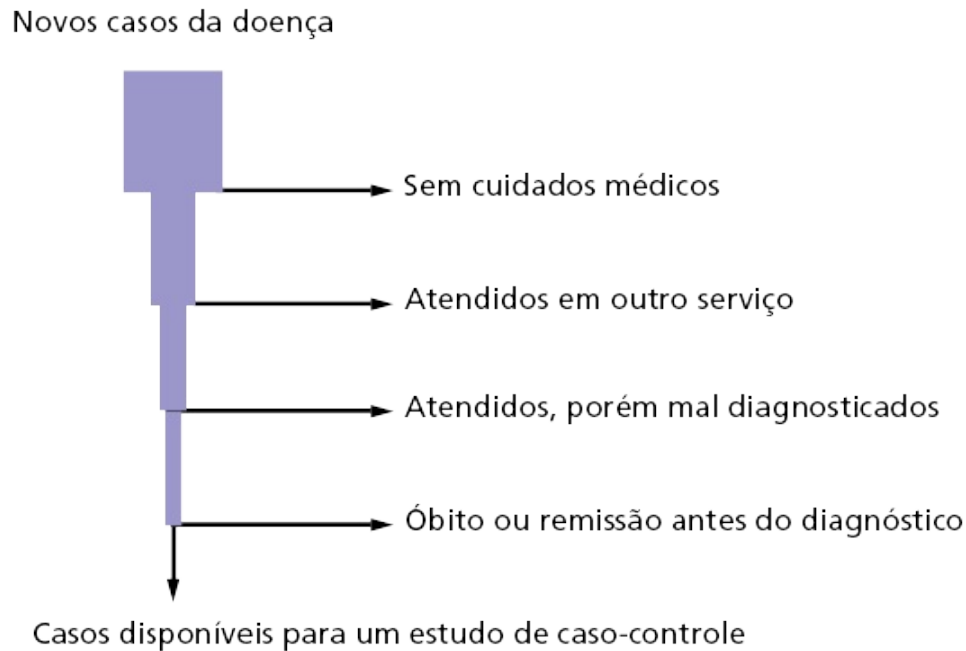



FIGURA 8.2 Motivos pelos quais os casos em um estudo de caso-controle podem não ser representativos de todos os casos com a doença.

Em geral, o viés de amostragem torna-se preocupante quando a amostra de casos é não representativa com relação ao fator de risco sob estudo. Doenças que normalmente exigem hospitalização e que são de fácil diagnóstico, como fratura de bacia e amputações traumáticas, podem ser amostradas seguramente a partir dos casos diagnosticados e acessíveis, pelo menos em países desenvolvidos. Por outro lado, condições clínicas que podem não chegar ao cuidado médico não são adequadas para esses estudos retrospectivos devido à seletividade que precede o diagnóstico. Por exemplo, mulheres atendidas em um ambulatório de ginecologia com aborto espontâneo no primeiro trimestre provavelmente diferem do total da população de mulheres com aborto espontâneo, muitas das quais não buscam atendimento médico. Assim, mulheres com história prévia de infertilidade seriam hiper-representadas em uma amostra de ambulatório, aquelas com baixo acesso aos cuidados pré-natais seriam sub-representadas. Se uma variável preditora de interesse está associada ao cuidado ginecológico na população (como uso de dispositivo intrauterino [DIU] no passado), amostrar os casos a partir de ambulatório de ginecologia poderia ser fonte importante de viés. Se, por outro lado, o preditor não está relacionado ao cuidado ginecológico (como no caso do

tipo sanguíneo da mulher), diminui a possibilidade de uma amostra de ambulatório não ser representativa.

Embora seja importante refletir sobre essas questões, a seleção dos casos geralmente fica limitada às fontes acessíveis de sujeitos. Mesmo que a amostra dos casos não seja completamente representativa, ela pode ser tudo de que se dispõe para trabalhar. As decisões mais difíceis do delineamento de um estudo de caso-controle, portanto, costumam estar relacionadas à tarefa mais aberta de selecionar os controles apropriados. A meta geral é amostrar os controles a partir da população que teria se tornado um caso no estudo se tivesse desenvolvido a doença. A seguir, apresentamos quatro estratégias para a amostragem de controles:

- **Controles ambulatoriais ou hospitalares.** Uma estratégia para compensar o possível viés de seleção causado pela obtenção dos casos em ambulatório ou hospital é selecionar os controles no mesmo local. Por exemplo, em um estudo sobre o uso de DIU no passado como fator de risco para aborto espontâneo, os controles poderiam ser amostrados de uma população de mulheres que buscam cuidado médico para outras queixas (p. ex., vaginite) no mesmo ambulatório de ginecologia. Presume-se que esses controles, se comparados a uma amostra aleatória de mulheres da mesma região, representam melhor a população de mulheres que, se tivessem desenvolvido aborto espontâneo, teriam ido à clínica e se tornado um caso no estudo.

  No entanto, selecionar uma amostra não representativa de controles para compensar por uma amostra não representativa de casos pode ser problemático. Se o fator de risco de interesse também é causa das doenças que levam os controles a buscar atendimento médico, a prevalência do fator de risco no grupo-controle seria falsamente elevada, diminuindo ou revertendo a associação entre o fator de risco e o desfecho. Se, por exemplo, muitas mulheres no grupo-controle tivessem procurado atendimento no ambulatório devido a uma condição médica associada com o uso prévio de DIU (p. ex., infertilidade causada por modelos antigos de DIU), haveria excesso de usuárias de DIU entre os controles, reduzindo a magnitude da associação entre uso prévio de DIU e aborto espontâneo no estudo.

  Como os sujeitos-controle de hospitais e ambulatórios

frequentemente apresentam problemas de saúde associados aos fatores de risco sob estudo, seu uso pode dar origem a achados enganadores. Por esse motivo, é fundamental refletir sobre se a conveniência de usar controles hospitalares ou ambulatoriais compensa as ameaças que eles impõem à validade do estudo.

- **Uso de amostras de casos e de controles de base populacional.** Graças à rápida expansão do uso de registros de doenças para populações geograficamente definidas e para usuários de planos de saúde, é possível hoje realizar estudos de caso-controle de base populacional para diversas patologias. Os casos obtidos desses registros são geralmente representativos da população geral de pacientes na área com a doença, o que simplifica a escolha de um grupo-controle: ele deve ser uma amostra representativa dos não casos na população coberta pelo registro. No Exemplo 8.1, todos os residentes da cidade haviam sido registrados pela prefeitura local, o que facilitou o processo de amostragem.

  Quando há registros disponíveis, os estudos de caso-controle de base populacional são claramente a melhor escolha. À medida que o registro de doenças se torna mais completo e a população coberta se aproxima da estabilidade (pouca migração), o estudo de caso-controle de base populacional aproxima-se de um estudo de caso-controle aninhado em um estudo de coorte ou em um ensaio clínico (ver mais adiante, neste capítulo, Estudos de caso-controle aninhados, estudos de caso-controle aninhados com amostragem por incidência-densidade e estudos de caso-coorte), pressupondo-se que os controles possam ser identificados e arrolados. Essas tarefas são relativamente simples quando a população foi enumerada e os registros estiverem disponíveis para os investigadores, como ocorreu no estudo sobre vitamina K e leucemia descrito no Exemplo 8.1. Quando esses registros não estão disponíveis, uma abordagem frequentemente utilizada é o método da discagem de dígito aleatório para telefones fixos com prefixos da região coberta pelo registro. (Quando os controles são selecionados dessa forma, os casos que não têm telefone fixo devem ser excluídos). Com o número crescente de famílias que optam por ter apenas telefone celular, essa abordagem tem se tornado problemática (12). É possível fazer discagem

de dígito aleatório incluindo números de telefone celular, mas isso deve ser feito com cuidado, encerrando a ligação imediatamente se o paciente estiver dirigindo e evitando fazer ligações pelas quais o paciente poderá ser cobrado.

É importante reconhecer, entretanto, que se pode introduzir viés sempre que os sujeitos precisarem ser contatados para obter informações, visto que alguns sujeitos (p. ex., aqueles que não falam inglês ou que têm deficiência auditiva) terão menor probabilidade de serem incluídos. Um problema semelhante pode ocorrer sempre que for necessário o consentimento informado.

- **Uso de dois ou mais grupos-controle.** A seleção de um grupo-controle pode ser traiçoeira, especialmente quando os casos não são uma amostra representativa dos indivíduos que têm a doença. Portanto, é às vezes recomendável que se usem dois ou mais grupos-controle selecionados de diferentes formas. O estudo do Public Health Service sobre a síndrome de Reye e sua associação ao uso de medicamentos (14), por exemplo, usou quatro tipos de controles: controles de serviços de emergência (atendidos na mesma emergência que o caso), controles hospitalares (internados no mesmo hospital que o caso), controles escolares (que frequentam a mesma escola ou creche que o caso) e controles comunitários (identificados por discagem aleatória). As razões de chances para o uso de salicilato obtidas da comparação de casos com cada um dos grupos-controle foram de pelo menos 30, todas elas com alta significância estatística. Achados consistentes de forte associação usando diferentes grupos-controle com diferentes tipos de viés de amostragem tornam mais convincente a inferência de que há uma associação real na população.

  Infelizmente, poucas associações têm razões de chances tão elevadas assim, e os vieses associados com diferentes estratégias de seleção de controles podem tornar conflitantes os resultados produzidos pelos diferentes grupos controles, o que revela a fragilidade inerente do delineamento de caso-controle para abordar a questão de pesquisa em pauta. Quando isso ocorre, o investigador deve buscar informações adicionais (p. ex., a queixa principal dos controles ambulatoriais) para buscar determinar a magnitude dos potenciais vieses de cada grupo

controle (Capítulo 9). De qualquer forma, é melhor ter resultados inconsistentes e concluir que não se sabe a resposta do que ter apenas um grupo-controle e chegar à conclusão errada.

- **Pareamento.** O pareamento é um método simples de garantir que casos e controles sejam comparáveis em relação a fatores importantes que estão relacionados à doença, mas que, neste estudo, não sejam de interesse para o investigador. Por exemplo, há tantos fatores de risco e doenças relacionados à idade e ao sexo que os resultados de um estudo podem não ser convincentes, a não ser que casos e controles sejam comparáveis quanto a essas duas variáveis. Uma forma de evitar esse problema é escolher controles que formem par com os casos em relação a essas variáveis preditoras constitucionais. O pareamento tem, no entanto, desvantagens importantes, especialmente quando preditores que podem mudar, como renda ou nível de colesterol sérico, são pareados. Os motivos para isso e as alternativas que muitas vezes são preferíveis ao pareamento serão discutidos no Capítulo 9.

Viés de medição diferencial e como controlá-lo

A segunda limitação importante dos estudos de caso-controle é o risco de viés devido ao **erro de aferição**. Ele é causado pela estratégia retrospectiva de medição das variáveis preditoras: tanto os casos quanto os controles devem recordar de exposições que ocorreram anos antes. Infelizmente, a memória das pessoas para exposições que ocorreram no passado está longe de ser perfeita. Se ela é igualmente imperfeita nos casos e nos controles, o problema é chamado de **erro de classificação não diferencial** da exposição, que torna mais difícil encontrar associações. (Em termos epidemiológicos, a razão de chances é modificada para tender a 1.) Mais preocupante, no entanto, é quando ser diagnosticado com uma doença leva os casos a lembrar ou relatar suas exposições de forma diferente em relação aos controles; esse **erro de classificação diferencial** da exposição, denominado **viés recordatório**, tem efeitos imprevisíveis sobre as associações medidas no estudo.

Por exemplo, a ampla publicidade sobre a relação entre exposição solar e melanoma maligno pode fazer com que os casos diagnosticados com esse tipo de câncer lembrem de sua exposição solar de forma diferente

quando comparados aos controles. Cockburn et al. (15) encontraram evidências para isso em um estudo inteligente que avaliou gêmeos discordantes em relação ao melanoma: a razão de chances pareada para ter tomado banho de sol na infância foi de 2,2 (IC 95%, 1,0 a 4,7) quando o gêmeo com melanoma foi questionado sobre qual dos dois irmãos tinha tomado mais banho de sol, mas foi de apenas 0,8 (0,4 a 1,8) quando o gêmeo sem melanoma respondeu à mesma questão. Entretanto, para algumas outras questões, como qual gêmeo se bronzeou mais ou teve mais queimadura solar, não houve evidência de viés recordatório.

O viés recordatório não ocorre no estudo de coorte, pois as perguntas sobre as exposições são feitas antes do diagnóstico da doença. Um estudo de caso-controle sobre melanoma maligno aninhado em uma coorte que coletou dados sobre exposição solar anos antes forneceu um teste direto do viés recordatório: os investigadores compararam a exposição solar autorrelatada nos casos e nos controles antes e depois de o caso ter sido diagnosticado com melanoma (16). Os investigadores encontraram algumas inacurácias na recordação da exposição, tanto nos casos quanto nos controles, mas pouca evidência de viés recordatório (16). Portanto, embora seja importante considerar a possibilidade de viés recordatório, esse viés não é necessariamente inevitável (17).

Além das estratégias para controlar viés de aferição apresentadas no Capítulo 4 (padronização das definições operacionais das variáveis, escolha de abordagens objetivas, suplementação das variáveis principais com dados de várias fontes, etc.), duas estratégias específicas podem ser usadas para evitar vieses na medição das exposições em estudos de caso-controle:

- **Usar dados registrados antes da ocorrência do desfecho.** Pode ser possível, por exemplo, avaliar registros médicos perinatais em um estudo de caso-controle sobre a administração IM de vitamina K como fator de risco para câncer. Essa estratégia, embora excelente, depende da disponibilidade de informações registradas sobre o fator de risco de interesse que sejam de razoável confiabilidade. Por exemplo, os dados sobre a administração de vitamina K muitas vezes estavam ausentes nos registros médicos, e a forma de lidar com esses dados faltantes afetou os resultados de alguns estudos sobre vitamina K como fator de risco

para câncer (8).

- **Usar cegamento (mascaramento).** A estratégia geral para o cegamento foi discutida no Capítulo 4, mas há algumas questões que são específicas ao delineamento de entrevistas em estudos de caso-controle. Teoricamente, é possível cegar tanto os observadores quanto os sujeitos do estudo para o estado de caso ou controle de cada sujeito e para o fator de risco que está sendo estudado; portanto, há quatro tipos possíveis de cegamento (Tabela 8.1).

TABELA 8.1 Estratégias para o cegamento em um estudo de caso-controle

INDIVÍDUO CEGADO	CEGAMENTO PARA O ESTADO DE CASO-CONTROLE	CEGAMENTO PARA A MEDIDA DO FATOR DE RISCO
Sujeito	É possível se tanto os casos quanto os controles tiverem doenças que possam estar relacionadas ao fator de risco.	Incluir fatores de risco falsos, para despistar, e suspeitar quando eles diferirem entre casos e controles. O cegamento pode não ser possível quando o fator de risco para a doença já for de conhecimento público.
Observador	É possível se os casos não puderem ser distinguidos dos controles pela aparência. Porém, sinais sutis e declarações não previstas dos sujeitos podem quebrar o cegamento alcançado.	É possível se o entrevistador não for o investigador, mas pode ser difícil de ser mantido.

O ideal seria que nem os sujeitos do estudo nem os investigadores soubessem quem é caso e quem é controle. Porém, na prática, isso é muito difícil, pois os sujeitos sabem quando estão doentes ou sadios. Assim, somente é possível mascarar o estado de caso-controle quando os controles são selecionados entre os pacientes que também apresentam doenças que acreditam estarem relacionadas aos fatores de risco em estudo. Os esforços de cegar os entrevistadores são prejudicados pela natureza óbvia de algumas doenças (é difícil um entrevistador não notar se o paciente tem icterícia ou se foi submetido à laringectomia) e pelos indícios percebidos nas respostas do paciente.

Em geral, é mais fácil cegar para o fator de risco em estudo do que para o estado de caso-controle. Os estudos de caso-controle muitas vezes são a primeira etapa na investigação de uma doença, portanto pode não haver

um fator de risco de interesse especial. Dessa forma, os sujeitos do estudo e os entrevistadores podem ser mantidos alheios às hipóteses de pesquisa pela inclusão de questões sobre fatores de risco plausíveis, porém não associados à doença, com o objetivo de despistá-los. Por exemplo, para testar a hipótese sobre se o consumo de mel estaria associado a um aumento do risco de botulismo infantil, poderiam ser incluídas na entrevista questões igualmente detalhadas sobre o consumo de iogurte e bananas. Esse tipo de cegamento não previne o viés diferencial, mas permite uma estimativa da ocorrência desse viés. Se os casos relatarem maior exposição ao mel, mas não houver aumento no consumo de outros alimentos, a possibilidade de viés de medição diferencial ficará reduzida. Essa estratégia não funcionaria se a associação entre botulismo infantil e mel já fosse de conhecimento público ou se os fatores de risco para despistar se mostrassem fatores de risco verdadeiros.

Cegar o observador para o estado de caso-controle é uma estratégia excelente para **medições laboratoriais**, como exames de sangue, e para radiografias. O cegamento, nessas circunstâncias, é de fácil aplicação e deve sempre ser usado. Alguém que não seja o indivíduo que fará a medição deve rotular cada amostra com um código de identificação. A importância do cegamento é ilustrada por 15 estudos de caso-controle que compararam medidas de massa óssea entre pacientes com fratura da bacia e controles; foram encontradas diferenças muito maiores nos estudos que usaram medições não cegas do que nos estudos cegos (18).

ESTUDOS DE CASO-CONTROLE ANINHADOS, ESTUDOS DE CASO-CONTROLE ANINHADOS COM AMOSTRAGEM POR INCIDÊNCIA-DENSIDADE E ESTUDOS DE CASO-COORTE

O delineamento do tipo **caso-controle aninhado** é basicamente um estudo de caso-controle que está “aninhado” em uma coorte (Figura 8.3). Muitas vezes, essa coorte já foi definida previamente pelo investigador como parte de um estudo de coorte formal no qual foram armazenadas amostras biológicas ou exames de imagem para análise futura, após a ocorrência dos desfechos. Outras vezes, o estudo de caso-controle é delineado do zero, aninhado em uma coorte ainda não definida, sendo o primeiro passo definir os casos na coorte.

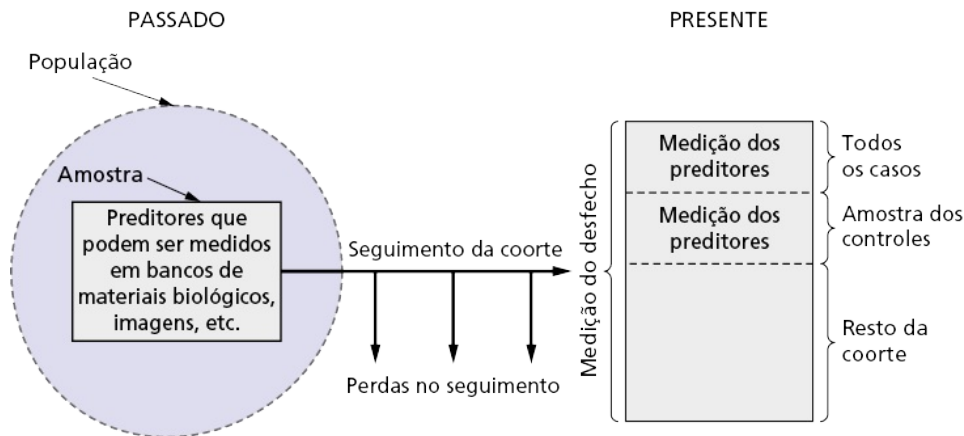



FIGURA 8.3 Um estudo de caso-controle aninhado pode ser prospectivo ou retrospectivo. Para a versão retrospectiva, os passos são:

- Identificar uma coorte com amostras, imagens e outros dados previamente armazenados.
- Medir a variável de desfecho que distingue casos de controles.
- Medir variáveis preditoras em materiais biológicos, imagens e outros dados armazenados desde que a coorte foi formada, assim como outras variáveis, em todos os casos e em uma amostra dos não-casos (controles).

Primeiro identifica-se uma coorte de sujeitos em risco para o desfecho cujo tamanho permita produzir um número suficiente de casos para responder à questão de pesquisa. Deve ser possível também medir a variável de exposição, seja porque amostras biológicas foram armazenadas ou porque registros médicos com informações sobre a exposição (ou os próprios sujeitos) estão disponíveis. Como descrito no Capítulo 7, a definição da coorte inclui os critérios de inclusão e exclusão que definem uma população em risco. Além disso, para cada sujeito, deve estar clara a **data de ingresso** na coorte. Ela pode ser uma data fixa (p. ex., todas as pessoas que atendiam os critérios de inclusão e que estavam vinculadas a um plano de saúde em 1º de janeiro de 2008) ou pode ser uma data variável, na qual um período de risco inicia (p. ex., data da inclusão em um estudo de coorte ou data do primeiro infarto do miocárdio em um estudo sobre fatores de risco para infarto do miocárdio recorrente).

O investigador, então, descreve os critérios que definem o desfecho de interesse, que sempre terá ocorrido após a data de ingresso na coorte e

antes do final do período de seguimento. Se o desfecho for raro, o seguimento estiver quase completo e uma única aferição da exposição na linha de base for suficiente, então o procedimento é simples. O investigador identifica todos os indivíduos na coorte que desenvolveram o desfecho até o final do período de seguimento (os casos) e então seleciona uma amostra aleatória dos sujeitos que também eram parte da coorte, mas não desenvolveram o desfecho (os controles). O investigador então mede as variáveis preditoras para os casos e para os controles e compara os níveis do fator de risco nos casos aos níveis na amostra de controles. Esse é o delineamento de caso-controle aninhado simples (Exemplo 8.2).

EXEMPLO 8.2 Delineamento de caso-controle aninhado simples

Para determinar se níveis elevados de hormônios sexuais estão associados a um aumento do risco de câncer de mama, Cauley e colaboradores (19) conduziram um estudo de caso-controle aninhado. As etapas básicas da realização do estudo foram:

1. **Identificação de uma coorte.** Os investigadores usaram a coorte do Study of Osteoporotic Fractures (SOF). Essa foi uma boa escolha, pois amostras de soro de membros da coorte haviam sido coletadas pelos investigadores na linha de base e armazenadas a uma temperatura de –190ºC, com a expectativa de que um estudo como este seria delineado.
2. **Identificação dos casos no final do período de seguimento.** Com base nas respostas aos questionários de seguimento e na revisão das certidões de óbito, os investigadores identificaram 97 sujeitos que desenvolveram uma primeira manifestação de câncer de mama durante 3,2 anos de seguimento.
3. **Seleção dos controles.** Os investigadores selecionaram uma amostra aleatória de 244 mulheres da coorte que não desenvolveram câncer de mama durante esse período.
4. **Medição dos preditores.** Os níveis de hormônios sexuais, incluindo estradiol e testosterona, foram medidos nas amostras de soro congeladas do exame da linha de base dos casos e controles. O procedimento laboratorial foi cegado quanto à procedência das

amostras, isto é, se eram dos casos ou dos controles.

Mulheres com níveis elevados de estradiol ou testosterona apresentaram um risco três vezes maior de vir a ter um diagnóstico subsequente de câncer de mama do que mulheres com níveis muito baixos desses hormônios.

Quando o seguimento é variável ou **incompleto**, ou a exposição **varia com o tempo**, não basta fazer uma única aferição da exposição no momento do ingresso na coorte nos casos e em uma amostra aleatória dos controles. Nessas situações, é melhor delinear um **estudo de caso-controle com amostragem por** *incidência-densidade* e amostrar os controles a partir de **conjuntos em risco**. Os conjuntos em risco são definidos para cada caso, no momento em que ele ocorre, como os membros da coorte que foram acompanhados pelo mesmo período de tempo, mas ainda não se tornaram casos (Figura 8.4). Assim como ocorre com qualquer outro tipo de pareamento de casos com controles, esse pareamento de acordo com o tempo de seguimento deve ser levado em conta na hora de fazer as análises estatísticas.

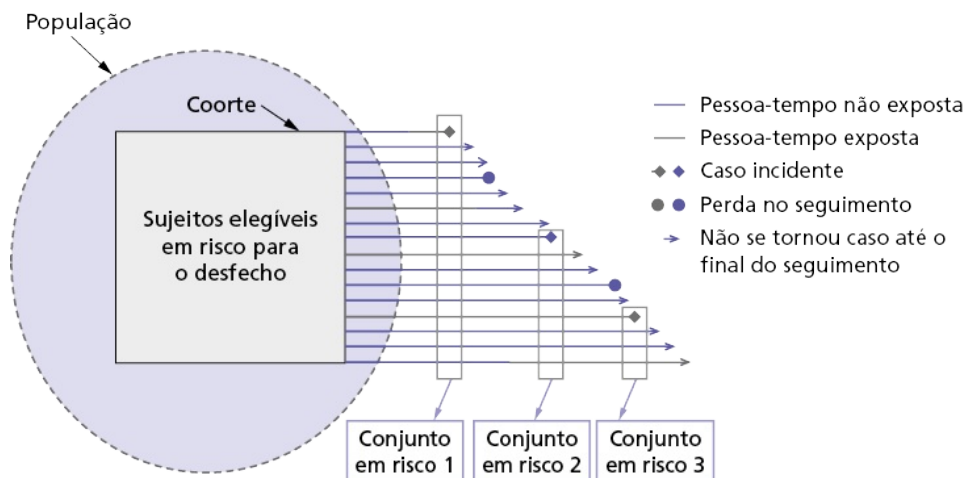



FIGURA 8.4 Um estudo de caso-controle com amostragem por incidência densidade pode ser prospectivo ou retrospectivo. Na versão prospectiva, os passos são:

- Definir os critérios de seleção e recrutar uma coorte da população.
- Definir a data de entrada para cada membro da coorte de modo a alinhar os tempos de seguimento. Armazenar amostras biológicas, imagens, etc., para análise posterior.

- Seguir a coorte para identificar os casos e a data em que foram diagnosticados.
- Amostrar um ou mais controles para cada caso a partir de "conjuntos em risco", definidos como membros da coorte que foram seguidos pelo mesmo período de tempo que o caso e não se tornaram caso, nem morreram ou foram perdidos no seguimento até o momento em que o caso foi diagnosticado.
- Medir variáveis preditoras em amostras biológicas, imagens, etc., armazenados desde a linha de base, assim como outras variáveis atuais, em casos e controles pareados.

Por exemplo, se o ingresso na coorte ocorreu em uma data fixa (p.ex., 1º de janeiro de 2008), os controles para um caso diagnosticado em 1º de julho de 2009 seriam amostrados a partir dos sujeitos que ainda não desenvolveram o desfecho até 1º de julho de 2009. Se a data de ingresso na coorte for variável, os controles para um caso diagnosticado 18 meses após o ingresso seriam amostrados a partir daqueles sujeitos que ainda não se tornaram casos após 18 meses de seguimento. Dependendo da hipótese de pesquisa do investigador, os valores da exposição no ingresso ou em algum outro momento após o ingresso poderiam ser comparados entre casos e controles.

Essa amostragem baseada em conjuntos em risco introduz a complexidade de que um sujeito pode ter sido selecionado inicialmente como controle para um caso que ocorreu no início do seguimento e depois ele próprio pode ter se tornado um caso, talvez após o seu valor para a variável de exposição ter mudado. Na prática, o que esse delineamento faz (com o auxílio da análise estatística apropriada) é considerar sequencialmente parcelas de pessoa-tempo em risco, usando, para cada parcela, os valores das variáveis preditoras para predizer a ocorrência de casos naquela parcela específica de pessoa-tempo, sendo que os limites de cada parcela são definidos pela ocorrência dos casos. Isso é denominado amostragem por **incidência-densidade** (Exemplo 8.3).

EXEMPLO 8.3 Delineamento de caso-controle aninhado com amostragem por incidência-densidade

Para investigar uma possível associação entre o antidiabético oral

pioglitazona e câncer de bexiga, investigadores de Montreal (20) realizaram um estudo de caso-controle aninhado ao United Kingdom General Practice Research Database, que contém registros médicos de atenção primária completos para mais de 10 milhões de pessoas vinculadas a mais de 600 clínicas de atenção primária no Reino Unido. As etapas do estudo foram:

1. **Identificação da coorte e do período de tempo em risco.** Os investigadores incluíram adultos que receberam pela primeira vez prescrição de antidiabéticos orais entre 1º de janeiro de 1988 e 31 de dezembro de 2009, que haviam sido acompanhados no banco de dados por pelo menos um ano antes da prescrição e que tinham pelo menos 40 anos de idade no momento da prescrição. A data dessa primeira prescrição de antidiabético foi a data de ingresso na coorte. Os participantes foram acompanhados até o momento do diagnóstico de câncer de bexiga, de morte por qualquer causa, do fim da vinculação à clínica de atenção primária ou do término do estudo em 31 de dezembro de 2009, o que ocorreu primeiro. Os sujeitos com história prévia de câncer de bexiga foram excluídos.
2. **Identificação dos casos, incluindo as datas de ocorrência.** Os investigadores identificaram os casos incidentes de câncer de bexiga por meio do "*Read codes*", um sistema para codificação dos diagnósticos validado na base de dados de pesquisa em atenção primária (21). Para levar em consideração a expectativa de que o efeito da pioglitazona sobre o risco de câncer não seria imediato, eles excluíram os casos que ocorreram no primeiro ano após o ingresso na coorte. Assim, identificaram os 376 casos restantes de câncer de bexiga.
3. **Amostragem dos controles a partir de "conjuntos em risco" pareados a cada caso.** Os investigadores sortearam até 20 controles para cada caso que não haviam sido diagnosticados com câncer de bexiga até o momento do diagnóstico do caso. Os controles foram pareados por ano de nascimento, ano de ingresso na coorte, sexo e duração do seguimento. O número total de controles pareados foi de 6.699 (número médio de controles por caso = 17,8).[1]

4. **Definição e medição dos preditores.** O preditor principal de interesse foi receber uma prescrição de pioglitazona ou rosiglitazona, outro antidiabético da mesma classe da pioglitazona. A prescrição deveria ter ocorrido pelo menos um ano antes do diagnóstico do caso no conjunto de risco. Quatro níveis de exposição foram definidos: prescrição apenas de pioglitazona, prescrição apenas de rosiglitazona, prescrição de ambos, ou prescrição de nenhum.

Os autores usaram (apropriadamente) a regressão logística condicional para analisar os dados; essa estratégia leva em conta a natureza pareada dos dados e, devido à amostragem por conjuntos em risco, permite estimar as razões de taxas ajustadas (22). Eles encontraram razões de taxas ajustadas de 1,83 (IC 95% 1,10 a 3,05) para uso exclusivo de pioglitazona, 1,14 (IC 95% 0,78 a 1,68) para uso exclusivo de rosiglitazona e 0,78 (IC 95% 0,18 a 3,29) para uso de ambos. (O IC maior no último grupo reflete um tamanho de amostra muito menor [N = 2 casos e 56 controles.]) Também encontraram evidência de relação dose-resposta entre uso de pioglitazona e câncer de bexiga: a razão de taxas ajustada para a dose cumulativa de 28 gramas ou mais foi de 2,54 (1,05-6,14), *P* para tendência de dose-respostα = 0,03.

O delineamento de **caso-coorte *aninhado*** é semelhante ao delineamento de caso-controle aninhado simples, exceto pelo fato de que, em vez de selecionar controles que não desenvolveram o desfecho de interesse, o investigador seleciona uma amostra aleatória de todos os membros da coorte, independentemente dos desfechos. Alguns sujeitos que fazem parte da amostra aleatória poderão ter desenvolvido o desfecho (o número é muito pequeno quando o desfecho é pouco comum). Uma vantagem do delineamento de caso-coorte é que uma única amostra aleatória da coorte pode fornecer os controles para diversos estudos de caso-controle sobre diferentes desfechos. Além disso, a amostra aleatória da coorte fornece informações sobre a prevalência geral de fatores de risco na coorte.

Pontos fortes

Os estudos de **caso-controle aninhados** e os de **caso-coorte** são especialmente úteis para medições caras de soro ou outras amostras biológicas ou de imagens que foram arquivadas no início do estudo e preservadas para análise posterior. Fazer aferições caras em todos os casos e em uma amostra dos controles resulta em um custo muito menor do que fazer as aferições em toda a coorte.

Esse delineamento preserva todas as vantagens dos estudos de coorte decorrentes da medição das variáveis preditoras antes da ocorrência dos desfechos. Além disso, evita os potenciais vieses dos estudos de caso-controle convencionais que não podem fazer aferições em casos fatais e que selecionam casos e controles de populações diferentes.

Pontos fracos

Esses delineamentos compartilham certas desvantagens com outros delineamentos observacionais: a possibilidade de as associações observadas terem ocorrido devido a variáveis confundidoras não aferidas ou aferidas de forma imprecisa e de as aferições na linha de base serem afetadas por uma doença pré-clínica assintomática.

Outras considerações

Os delineamentos de caso-controle aninhados e de caso-coorte têm sido usados com menor frequência do que deveriam. Ao planejar um estudo prospectivo de grande porte, deve-se sempre considerar preservar amostras biológicas (p. ex., bancos de soro congelado) ou armazenar imagens ou registros de análise dispendiosa para análises posteriores com estudos de caso-controle aninhados. Para tanto, deve-se garantir que as condições de armazenamento sejam capazes de preservar as substâncias de interesse durante muitos anos. Pode ser útil também coletar novas amostras ou informações, durante o período de seguimento, que poderão ser usadas nas comparações de casos e controles.

ESTUDOS CRUZADOS DE CASOS (*CASE-CROSSOVER STUDIES*)

O delineamento cruzado de casos (*case-crossover design*) é uma variante do delineamento de caso-controle, útil para estudar os efeitos a curto

prazo de exposições intermitentes. Assim como ocorre com os estudos de caso-controle tradicionais, esses são estudos retrospectivos que iniciam com um grupo de casos: indivíduos que desenvolveram o desfecho de interesse. No entanto, ao contrário dos estudos de caso-controle tradicionais, nos quais as exposições dos casos são comparadas com as exposições de um grupo de controles, em estudos cruzados de casos, cada caso serve como seu próprio controle. As exposições dos casos no momento do desfecho (ou logo antes) são comparadas com as exposições desses mesmos indivíduos em um ou mais momentos no tempo.

Por exemplo, McEvoy e colaboradores (23) estudaram vítimas de acidentes automobilísticos que relataram possuir ou ter usado um telefone celular. A partir dos registros de companhias telefônicas, eles compararam o uso de telefones celulares nos 10 minutos que antecederam o acidente com o uso dos celulares quando os sujeitos estavam dirigindo no mesmo horário do dia, 24 horas, 72 horas e 7 dias antes. O uso de telefones celulares foi mais frequente nos 10 minutos anteriores ao acidente do que nos períodos de tempo de comparação, com uma razão de chances de aproximadamente 4. A análise de um estudo cruzado de casos é semelhante àquela de um estudo de caso-controle pareado, exceto pelo fato de que as exposições dos controles são exposições nos casos em períodos de tempo diferentes, em vez de exposições em controles pareados. Isso é ilustrado no Apêndice 8A, cenário número 4. Delineamentos cruzados de casos têm sido utilizados em grandes populações para estudar exposições que variam ao longo do tempo, como os níveis de poluição atmosférica; para essa exposição, foram demonstradas associações com infarto do miocárdio (24, 25), consultas em serviços de emergência por doenças respiratórias (26) e até mesmo mortalidade infantil (27).

ESCOLHENDO ENTRE OS DELINEAMENTOS OBSERVACIONAIS

Os prós e os contras dos principais delineamentos observacionais apresentados nos últimos dois capítulos são resumidos na Tabela 8.2. Já descrevemos essas questões detalhadamente, portanto apenas faremos aqui uma observação final. Nenhum desses delineamentos é melhor ou

pior do que os demais; cada um tem seu lugar definido e serve a um determinado propósito, dependendo da questão de pesquisa e das circunstâncias envolvidas.

TABELA 8.2 Vantagens e desvantagens dos principais delineamentos observacionais

DELINEAMENTO	VANTAGENS	DESVANTAGENS*
	Transversal	
	Duração relativamente curta	Não estabelece uma sequência de eventos
	Um bom primeiro passo para um estudo de coorte	Não é factível para preditores ou desfechos raros
	Produz prevalências e razões de prevalências	Não produz incidência
	Coorte	
Todos	Estabelecem a sequência dos eventos	Costumam exigir grandes tamanhos de amostra
	Permitem estudar vários preditores e desfechos	Menos factíveis para desfechos raros
	O número de eventos dos desfechos cresce com o tempo	
	Produzem incidência, risco relativo e excesso de risco	
Prospectiva	Maior controle sobre a seleção dos sujeitos e sobre as aferições	O seguimento pode ser longo
	Evita vieses na medição dos preditores	Muitas vezes tem custo elevado
Retrospectiva	O seguimento ocorreu no passado	Menor controle sobre a seleção dos sujeitos e sobre as aferições
	Menor custo	
Coortes múltiplas	Úteis quando coortes diferentes têm exposições diferentes ou raras	Potencial para vieses e confundimento na amostragem de várias populações
	Caso-controle	
	Útil para desfechos raros	Vieses e confundimento por amostrar duas populações
	Curta duração, pequeno tamanho de amostra	Viés de aferição diferencial
	Custo relativamente baixo	Limitado a uma única variável de desfecho
	Produz razão de chances (geralmente uma boa aproximação do risco relativo para desfechos incomuns)	Não estabelece uma sequência clara de eventos
		Não produz prevalência, incidência ou excesso de risco
	Delineamentos híbridos	
Caso-controle aninhado	Mesmas vantagens que um delineamento de coorte retrospectiva, mas muito mais eficiente	Podem não estar disponíveis coortes ou materiais biológicos armazenados

Caso-controle aninhado com amostragem por incidência-densidade	Permite analisar as relações de risco levando em consideração mudanças ao longo do tempo nos níveis dos fatores de risco e as perdas no seguimento	Requer a aferição dos níveis dos fatores de risco e da incidência de casos ao longo do tempo durante o seguimento; geralmente reque uma coorte definida preexistente
Caso-coorte aninhado	Pode usar um único grupo-controle para vários estudos	Podem não estar disponíveis coortes ou materiais biológicos armazenados
Estudo cruzado de casos	Os casos servem como seus próprios controles, reduzindo o erro aleatório e o confundimento	Requer circunstâncias especiais

* Todos esses delineamentos observacionais têm a desvantagem (quando comparados aos ensaios clínicos randomizados) de serem suscetíveis à influência de variáveis confundidoras – ver Capítulo 9.

RESUMO

1. Em um **estudo de caso-controle**, a prevalência de um fator de risco em uma amostra de sujeitos que têm o desfecho de interesse (**os casos**) é comparada com a prevalência em uma amostra que não tem (**os controles**). Esse delineamento, no qual as pessoas com e sem a doença são amostradas separadamente, tem **custo relativamente baixo** e é extremamente **eficiente** para estudar **doenças raras**.
2. Um problema com os estudos de caso-controle é sua suscetibilidade ao **viés de amostragem**. Quatro abordagens para reduzir o viés de amostragem são (a) amostrar os controles e os casos da **mesma forma** (conscientemente não representativa); (b) realizar um estudo de base populacional; (c) usar **vários grupos controle**, amostrados de formas diferentes; e (d) **parear** os casos e controles.
3. O outro problema importante com os estudos de caso-controle é o delineamento retrospectivo, que os torna suscetíveis ao **viés de aferição** que afeta casos e controles de forma diferente. Esse viés pode ser reduzido fazendo as **aferições da variável preditora antes da ocorrência do desfecho** e **cegando** os sujeitos e os observadores.
4. A melhor forma de **evitar tanto o viés de amostragem quanto o de aferição** é delinear um **estudo de caso-controle aninhado**, no qual as amostras de casos e controles são sorteadas de um estudo de coorte maior após sua conclusão. Além de controlar esses dois tipos de viés, esse delineamento permite fazer, no final do estudo, em um **número relativamente pequeno de sujeitos,** medições caras em amostras de

soro, exames de imagem, e assim por diante, coletados na linha de base.

5. O **delineamento de caso-controle aninhado com amostragem por incidência-densidade** permite analisar relações de risco, levando em consideração **mudanças ao longo do tempo** nos níveis do **fator de risco** e na **disponibilidade do seguimento**.
6. O **delineamento de caso-coorte aninhado** usa uma amostra aleatória de toda a coorte no lugar dos não casos; essa amostra pode servir como grupo controle para estudar **mais de um desfecho**, e fornece informações diretas sobre a prevalência geral de fatores de risco na coorte.
7. **Estudos cruzados de casos** são uma variação do delineamento de caso-controle pareado na qual as observações em dois ou mais momentos no tempo permitem que cada caso sirva como seu próprio controle.

APÊNDICE 8A
Calculando medidas de associação

1. **Estudo transversal.** Reijneveld (21) realizou um estudo transversal sobre o tabagismo materno como fator de risco para cólica em lactentes. Resultados parciais são apresentados abaixo:

TABELA 8A.1

	VARIÁVEL DE DESFECHO:		
VARIÁVEL PREDITORA	COM CÓLICA	SEM CÓLICA	TOTAL
Mãe fuma 15-20 cigarros/dia	15 (*a*)	167 (*b*)	182 (*a* + *b*)
Mãe não fuma	111 (c)	2.477 (*d*)	2.588 (*c* + *d*)
Total	126 (*a* + *c*)	2.644 (*b* + *d*)	2.770 (*a* + *b* + *c* + *d*)

Prevalência de cólica quando as mães fumam = *a*/(*a* + *b*) = 15/182 = 8,2%.
Prevalência de cólica quando as mães não fumam = *c*/(*c* + *d*) = 111/2.588 = 4,3%.
Prevalência geral de cólica = (*a* + *c*)/(*a* + *b* + *c* + *d*) = 126/2.770 = 4,5%.

$$\text{Razão de prevalências}^{2} = \frac{8,2\%}{4,3\%} = 1,9$$

$$\text{Excesso de prevalência}^{2} = 8,2\% - 4,3\% = 3,9\%$$

Em outras palavras, a cólica foi quase duas (1,9) vezes mais comum e ocorreu com uma frequência 4% maior em crianças cujas mães fumam.

2. **Estudos de caso-controle.** A questão de pesquisa do Exemplo 8.2 foi "Há associação entre vitamina K IM e o risco de leucemia infantil?". O estudo mostrou que 69/107 casos e 63/107 controles haviam recebido vitamina K IM. A seguir, apresentamos uma tabela 2 × 2 com esses achados:

TABELA 8A.2

	VARIÁVEL DE DESFECHO: DIAGNÓSTICO	
VARIÁVEL PREDITORA: ADMINISTRADA VITAMINA K INTRAMUSCULAR?	LEUCEMIA INFANTIL	CONTROLE
Sim	69 (a)	63 (b)
Não	38 (c)	44 (d)
Total	107	107

$$\text{Risco relativo} \approx \textit{razão de chances} = \frac{ad}{bc} = \frac{69 \times 44}{63 \times 38} = 1{,}27$$

Como a doença (neste caso, leucemia) é rara, a razão de chances fornece uma boa estimativa do risco relativo. Portanto, a leucemia teve uma probabilidade 1,3 vezes maior de ocorrer após a administração de vitamina K, mas isso não foi estatisticamente significativo.[3]

3. **Estudo de caso-controle pareado.**

(Para ilustrar a semelhança na análise de um estudo de caso-controle pareado e um estudo cruzado de casos, usaremos o mesmo exemplo para ambos.) A questão de pesquisa é se o uso de telefones celulares aumenta o risco de acidentes automobilísticos em indivíduos que possuem telefone celular. Um estudo tradicional de caso-controle pareado poderia considerar a frequência autorrelatada de uso de telefone celular ao dirigir como fator de risco. Assim, os casos seriam pessoas que sofreram acidentes, e eles poderiam ser comparados com controles não envolvidos em acidentes, pareados por idade, sexo e prefixo do telefone celular. Os casos e os controles seriam questionados sobre o uso de telefone celular enquanto dirigem. (Para simplificar, para fins deste exemplo, dicotomizamos a exposição e consideramos os indivíduos como "usuários" ou "não usuários" de telefones celulares ao dirigir.) Então classificamos cada par de casos e controles em: ambos usuários, nenhum usuário, caso usuário e controle não usuário ou controle usuário e caso não usuário. Para um tamanho de amostra de 300 pares, os resultados poderiam ser como os mostrados a seguir:

TABELA 8A.3

	CASOS (VÍTIMAS DE ACIDENTES AUTOMOBILÍSTICOS)		
CONTROLES PAREADOS	USUÁRIOS	NÃO USUÁRIOS	TOTAL
Usuários	110	40	150
Não usuários	90	60	150
Total	200	100	300

A Tabela 8A.3 mostra que em 90 dos pares o caso falou no telefone celular pelo menos uma vez enquanto dirigia (usuário), mas não o seu controle pareado, e em 40 dos pares o usuário foi o controle, e não o caso. Note que essa tabela 2 × 2 é diferente da tabela 2 × 2 do estudo não

pareado sobre vitamina K mencionado na questão 2, na qual o número dentro de cada casela representa o número de pacientes naquela casela. Na tabela para o estudo de caso-controle *pareado*, o número em cada casela corresponde ao número de *pares* de sujeitos naquela casela ; o *N* total na Tabela 8A.3, portanto, é 600 (300 casos e 300 controles). A razão de chances para uma tabela como essa é simplesmente a razão entre os dois tipos de pares discordantes; na Tabela 8A.3, a RC = 90/40 = 2,25. Isso implica que os usuários de telefones celulares tinham mais do que o dobro de chances de se envolverem em um acidente automobilístico.

4. **Estudo cruzado de casos.** Agora considere o estudo cruzado de casos sobre a mesma questão. Os dados do estudo de McEvoy et al. são apresentados abaixo.

TABELA 8A.4

	MOMENTO DO ACIDENTE		
SETE DIAS ANTES	FALOU NO CELULAR	NÃO FALOU NO CELULAR	TOTAL
Falou no celular	5	6	11
Não falou no celular	27	288	315
Total	32	294	326

Para o estudo cruzado de casos, cada casela na tabela corresponde a um número de sujeitos, e não de pares. Cada uma, porém, representa dois períodos no tempo para o mesmo sujeito: o período imediatamente anterior ao acidente e um período de comparação de sete dias antes. Portanto, o número 5 na casela superior à esquerda significa que 5 motoristas envolvidos em acidentes automobilísticos falaram no celular logo antes do acidente e também dirigiram falando no celular durante o período de comparação de sete dias antes. De forma semelhante, seis motoristas se envolveram em acidentes automobilísticos e não estavam falando no celular no momento do acidente, mas falaram no celular no período de tempo de comparação de sete dias antes. A RC é a razão entre os números de períodos de tempo discordantes, neste exemplo 27/6 = 4,5, significando que dirigir durante períodos de uso de telefone celular está associado com chances 4,5 vezes maiores de acidente automobilístico do que dirigir durante períodos sem uso de telefone celular.

APÊNDICE 8B

Por que a razão de chances pode ser usada como estimativa para o risco relativo em um estudo de caso-controle

Os dados em um estudo de caso-controle representam duas amostras: os casos são selecionados a partir de uma população de pessoas que têm a doença e os controles de uma população de pessoas que não a têm. Mede-se, então, a variável preditora (fator de risco), produzindo a tabela 2 × 2 apresentada a seguir:

	Casos	**Controles**
Fator de risco presente	*a*	*b*
Fator de risco ausente	*c*	*d*

Se essa tabela 2 × 2 representasse os dados de um estudo de coorte, então a incidência da doença naqueles com o fator de risco seria a/(a + b) e o risco relativo seria simplesmente [a/(a + b)] / [c/(c + d)]. No entanto, no estudo de caso-controle, não é correto computar a incidência ou o risco relativo dessa forma, pois as duas amostras não foram sorteadas da população nas mesmas proporções. Em geral, os números de casos e controles nas amostras são grosseiramente iguais; porém, na população, há muito menos casos que controles. Em situações como essa, o risco relativo pode ser estimado por meio da RC, isto é, o produto cruzado da tabela 2 × 2, ad/cb.

É difícil compreender isso intuitivamente, porém, é relativamente fácil demonstrar esse fato algebricamente. Considere a situação para a população total, representada por a', b', c' e d'.

	Com doença	**Sem doença**
Fator de risco presente	*a'*	*b'*
Fator de risco ausente	*c'*	*d'*

Aqui é correto calcular o risco de desenvolver a doença para indivíduos com o fator de risco usando a fórmula $a'/(a' + b')$; o risco de desenvolver a

doença para aqueles sem o fator de risco usando a fórmula $c'/(c' + d')$; e o risco relativo a partir da fórmula $[a'/(a' + b')]/[c'/(c' + d')]$. Como discutido, $a'/(a' + b')$ não é igual a $a/(a + b)$. No entanto, se a doença for relativamente pouco comum (como a maioria das doenças o são), a' é muito menor que b', e c' é muito menor que d'. Isso significa que $a'/(a' + b')$ pode ser aproximado por a'/b' e que $c'/(c' + d')$ pode ser aproximado por c'/d'. Portanto, o risco relativo para a população pode ser aproximado pela fórmula a seguir:

$$\frac{a'/(a' + b')}{c'/(c' + d')} \approx \frac{a'/b'}{c'/d'}$$

O último termo é a RC (*razão de chances*) na população (literalmente, a razão entre as chances [*odds*] de aqueles com o fator de risco desenvolverem a doença, a'/b', e as chances [*odds*] de aqueles sem o fator de risco desenvolverem a doença, c'/d'). Isso pode ser reordenado como o produto cruzado:

$$\left(\frac{a'}{c'}\right)\left(\frac{d'}{c'}\right) = \left(\frac{a'}{c'}\right)\left(\frac{d'}{b'}\right)$$

No entanto, se os casos forem representativos de todos os casos na população (isto é, têm a mesma prevalência do fator de risco), a'/c' na população é igual a a/c na amostra. Da mesma forma, se os controles forem representativos, b'/d' é igual a b/d.

Portanto, os parâmetros populacionais nesse último termo podem ser substituídos pelos parâmetros amostrais, resultando na constatação de que a razão de chances observada na amostra ad/bc é uma boa aproximação do risco relativo na população $[a'/(a' + b')]/[c'/(c' + d')]$, desde que a doença seja rara e o erro de amostragem (tanto o sistemático como o aleatório) seja pequeno.

REFERÊNCIAS

1. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971;284(15):878–881.
2. Beal SM, Finch CF. An overview of retrospective case–control studies investigating the relationship between prone sleeping position and SIDS. *J Paediatr Child Health* 1991;27(6):334–339.
3. Mitchell EA, Hutchison L, Stewart AW. The continuing decline in SIDS mortality.

Arch Dis Child 2007;92(7):625–626.
4. Golding J, Greenwood R, Birmingham K, Mott M. Childhood cancer, intramuscular vitamin K, and pethidine given during labour. *BMJ* 1992;305(6849):341–346.
5. Golding J, Paterson M, Kinlen LJ. Factors associated with childhood cancer in a national cohort study. *Br J Cancer* 1990;62(2):304–308.
6. von Kries R, Gobel U, Hachmeister A, et al. Vitamin K and childhood cancer: a population based case–control study in Lower Saxony, Germany. *BMJ* 1996;313(7051):199–203.
7. Fear NT, Roman E, Ansell P, et al. Vitamin K and childhood cancer: a report from the United Kingdom Childhood Cancer Study. *Br J Cancer* 2003;89(7):1228–1231.
8. Roman E, Fear NT, Ansell P, et al. Vitamin K and childhood cancer: analysis of individual patient data from six case-control studies. *Br J Cancer* 2002;86(1):63–69.
9. Kochen M, McCurdy S. Circumcision and the risk of cancer of the penis. A life-table analysis. *Am J Dis Child* 1980;134(5):484–486.
10. O'Brien KL, Selanikio JD, Hecdivert C, et al. Epidemic of pediatric deaths from acute renal failure caused by diethylene glycol poisoning. Acute Renal Failure Investigation Team. *JAMA* 1998; 279(15):1175–1180.
11. Fatal poisoning among young children from diethylene glycol-contaminated acetaminophen – Nigeria, 2008–2009. *MMWR Morb Mortal Wkly Rep* 2009;58(48):1345–1347.
12. Puumala SE, Spector LG, Robison LL, et al. Comparability and representativeness of control groups in a case–control study of infant leukemia: a report from the Children's Oncology Group. *Am J Epidemiol* 2009;170(3):379–387.
13. Voigt LF, Schwartz SM, Doody DR, et al. Feasibility of including cellular telephone numbers in random digit dialing for epidemiologic case–control studies. *Am J Epidemiol* 2011;173(1):118–126.
14. Hurwitz ES, Barrett MJ, Bregman D, et al. Public Health Service study of Reye's syndrome and medications. Report of the main study. *JAMA* 1987;257(14):1905–1911.
15. Cockburn M, Hamilton A, Mack T. Recall bias in self-reported melanoma risk factors. *Am J Epidemiol* 2001;153(10):1021–1026.
16. Parr CL, Hjartaker A, Laake P, et al. Recall bias in melanoma risk factors and measurement error effects: a nested case-control study within the Norwegian Women and Cancer Study. *Am J Epidemiol* 2009;169(3):257–266.
17. Gefeller O. Invited commentary: Recall bias in melanoma—much ado about almost nothing? *Am J Epidemiol* 2009;169(3):267–270; discussion 71–72.
18. Cummings SR. Are patients with hip fractures more osteoporotic? Review of the evidence. *Am J Med* 1985;78(3):487–494.
19. Cauley JA, Lucas FL, Kuller LH, et al. Elevated serum estradiol and testosterone concentrations are associated with a high risk for breast cancer. Study of Osteoporotic Fractures Research Group. *Ann Intern Med* 1999; 130(4 Pt 1):270–277.

20. Azoulay L, Yin H, Filion KB, et al. The use of pioglitazone and the risk of bladder cancer in people with type 2 diabetes: nested case–control study. *BMJ* 2012;344:e3645.

21. Hassey A, Gerrett D, Wilson A. A survey of validity and utility of electronic patient records in a general practice. *BMJ* 2001;322(7299):1401–1405.

22. Essebag V, Platt RW, Abrahamowicz M, et al. Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. *BMC Med Res Methodol* 2005;5(1):5.

23. McEvoy SP, Stevenson MR, McCartt AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ* 2005;331(7514):428.

24. Bhaskaran K, Hajat S, Armstrong B, et al. The effects of hourly differences in air pollution on the risk of myocardial infarction: case crossover analysis of the MINAP database. *BMJ* 2011;343:d5531.

25. Nuvolone D, Balzi D, Chini M, et al. Short-term association between ambient air pollution and risk of hospitalization for acute myocardial infarction: results of the cardiovascular risk and air pollution in Tuscany (RISCAT) study. *Am J Epidemiol* 2011;174(1):63–71.

26. Tramuto F, Cusimano R, Cerame G, et al. Urban air pollution and emergency room admissions for respiratory symptoms: a case-crossover study in Palermo, Italy. *Environ Health* 2011;10:31.

27. Scheers H, Mwalili SM, Faes C, et al. Does air pollution trigger infant mortality in Western Europe? A case-crossover study. *Environ Health Perspect* 2011;119(7):1017–1022.

28. Reijneveld SA, Brugman E, Hirasing RA. Infantile colic: maternal smoking as potential risk factor. *Arch Dis Child* 2000;83(4):302–303.

1Iremos mostrar, no Capítulo 9, que o ganho em poder ao amostrar mais de quatro controles por caso é pequeno, mas nesse caso o custo adicional era muito pequeno, pois os dados eletrônicos já estavam disponíveis. Mesmo com 20 controles por caso, a abordagem de caso-controle aninhada é muito mais eficiente em termos computacionais do que um estudo de coorte retrospectiva.

2 A razão de prevalências e o excesso de prevalência são os análogos transversais do risco relativo e do excesso de risco.

3 Os autores, na verdade, realizaram uma análise multivariada pareada, adequada ao delineamento, mas, neste caso, a razão de chances simples não pareada foi quase idêntica à relatada no estudo.

Fortalecendo a inferência causal em estudos observacionais

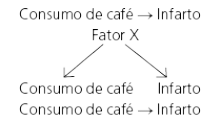
Thomas B. Newman, Warren S. Browner e Stephen B. Hulley

A maioria dos estudos observacionais tem como objetivo mostrar que uma variável preditora é causa de um desfecho. Por exemplo, que comer brócolis leva a uma diminuição do risco de câncer de cólon (algumas exceções são os estudos sobre testes diagnósticos e prognósticos, discutidos no Capítulo 12). Identificar associações causais entre uma variável preditora e um desfecho é importante, uma vez que ajuda a compreender a biologia de uma doença, a buscar formas de reduzir ou prevenir sua ocorrência e até mesmo a sugerir potenciais tratamentos. Entretanto, nem toda associação encontrada em um estudo observacional representa uma relação de causa-efeito.

De fato, há quatro outras possíveis explicações gerais para uma associação entre um preditor e um desfecho em um estudo observacional (Tabela 9.1). Duas delas, **acaso** e **viés**, dão origem a associações espúrias entre preditor e desfecho na amostra do estudo, que não existem na população.

TABELA 9.1 As cinco explicações possíveis para o fato de que tomar café está associado ao dobro do risco de infarto do miocárdio

EXPLICAÇÃO	TIPO DE ASSOCIAÇÃO	O QUE REALMENTE ESTÁ OCORRENDO NA POPULAÇÃO?	MODELO CAUSAL
1. Acaso (erro aleatório)	Espúria	O consumo de café e o infarto não estão relacionados	—
2. Viés (erro sistemático)	Espúria	O consumo de café e o infarto não estão relacionados	—
3. Efeito-causa	Real	O infarto é causa do consumo de café	

4. Confundimento	Real	Um terceiro fator leva ao consumo de café e causa o infarto	Consumo de café → Infarto Fator X ↙ ↘ Consumo de café Infarto Consumo de café → Infarto
5. Causa-efeito	Real	O consumo de café é causa do infarto	

As duas outras explicações, **efeito-causa** e **confundimento**, dão origem a associações que são reais na população, mas não são causais na direção de interesse. Para estabelecer que a relação causa-efeito é a explicação mais provável para uma associação, é preciso demonstrar que essas outras explicações são improváveis.

Costuma-se quantificar o efeito causal de uma variável preditora sobre um desfecho usando uma medida de associação, como o risco relativo (razão de riscos) ou a razão de chances. Por exemplo, suponha que um estudo mostre que tomar café está associado ao infarto do miocárdio, com um risco relativo de 2,0. Uma possibilidade – presumivelmente a que o investigador achou mais interessante – é que tomar café dobra o risco de sofrer um infarto. Antes de chegar a essa conclusão, no entanto, devem-se considerar e descartar as quatro explicações rivais.

Em virtude do acaso ou de viés, o consumo de café pode estar associado a um aumento de duas vezes no risco de infarto do miocárdio no estudo, mesmo essa associação não estando presente na população. Assim, o acaso e o viés dão origem a associações espúrias (isto é, não reais) em um estudo.

As outras duas alternativas – efeito-causa e confundimento – são fenômenos biológicos verdadeiros, o que significa que consumidores de café na população realmente têm o dobro do risco de infarto do miocárdio. Entretanto, esse aumento do risco não se dá por uma relação de causa-efeito. Em uma situação, a associação deve-se a uma relação efeito-causa: ter um infarto leva as pessoas a tomarem mais café. (Nesse caso, apenas causa e efeito ao contrário.) A última possibilidade, confundimento, ocorre quando um terceiro fator, como tipo de personalidade, leva tanto ao consumo de café quanto ao infarto do miocárdio.

No restante deste capítulo, iremos discutir estratégias para estimar e minimizar a possibilidade dessas quatro explicações alternativas para uma associação em um estudo observacional. Essas estratégias podem ser

usadas ao delinear um estudo ou ao analisar seus resultados. Embora o foco deste livro seja o delineamento da pesquisa, compreender as opções analíticas pode influenciar na escolha do delineamento Portanto, ambos os tópicos serão considerados neste capítulo.

ASSOCIAÇÕES ESPÚRIAS INDUZIDAS PELO ACASO

Imagine que não há associação verdadeira entre consumo de café e infarto em uma população, na qual 45% das pessoas tomam café. Se fôssemos selecionar 20 casos de infarto e 20 controles, esperaríamos que em torno de 9 pessoas em cada grupo (45% de 20) tomariam café.

Porém, pelo mero acaso, poderíamos sortear 12 consumidores de café entre os 20 casos de infarto, mas apenas 6 consumidores de café entre os 20 controles. Se isso ocorresse, observaríamos uma associação espúria entre consumo de café e infarto no nosso estudo.

O acaso é também denominado **erro aleatório**, uma vez que não tem explicação subjacente. Quando uma associação decorrente de erro aleatório é estatisticamente significativa, é denominada **erro tipo I** (Capítulo 5).

Existem estratégias para minimizar erros aleatórios, tanto na fase de delineamento quanto na fase de análise da pesquisa (Tabela 9.2). As estratégias de delineamento, como aumentar a **precisão das medidas** e o **tamanho de amostra**, foram discutidas nos Capítulos 4 e 6, respectivamente. A estratégia analítica de calcular **valores *P* e intervalos de confiança (**ICs) permite ao investigador quantificar a magnitude da associação observada em comparação com a que poderia resultar do mero acaso. Por exemplo, um valor *P* de 0,10 indica que o mero acaso poderia resultar em uma diferença de magnitude pelo menos semelhante à que os investigadores observaram em torno de 10% das vezes. ainda mais úteis que os valores *P*, os ICs mostram os valores possíveis para a estatística que descreve uma associação e que se enquadram na faixa de erro aleatório estimada no estudo.

TABELA 9.2 Como fortalecer a inferência de que uma associação representa causa-efeito: reduzindo e avaliando a probabilidade de associações espúrias

Tipo de Associação Espúria	Fase de Delineamento (Como Prevenir a Explicação Rival)	Fase de Análise (Como Avaliar a Explicação Rival)

Acaso (erro aleatório)	Aumentar o tamanho de amostra e outras estratégias para aumentar a precisão (Capítulos 4 e 6).	Calcular valores *P* e intervalos de confiança (ICs) e interpretá-los dentro do contexto de evidências anteriores (Capítulo 5).
Viés (erro sistemático)	Considerar cuidadosamente as potenciais consequências de cada diferença entre a questão de pesquisa e o plano de estudo (Figura 9.1); alterar o plano de estudo se necessário.	Verificar a consistência dos achados com estudos anteriores (dando preferência a estudos com metodologias diferentes).
	Coletar dados adicionais que permitirão avaliar a ocorrência de possíveis vieses.	Analisar dados adicionais para verifica a ocorrência de possíveis vieses.
	Não usar variáveis afetadas pelo preditor de interesse como critérios de inclusão ou como variáveis para pareamento.	Não controlar para variáveis afetadas pela sua variável preditora de interesse.

ASSOCIAÇÕES ESPÚRIAS INDUZIDAS POR VIÉS

Existem muitos tipos de viés, também denominado erro sistemático, e como lidar com eles é um tema importante deste livro. Às estratégias específicas descritas nos Capítulos 3, 4, 7 e 8, acrescentamos aqui uma abordagem geral para reduzir a probabilidade de viés.

Minimizando o viés

Como discutido no Capítulo 1, quase sempre há diferenças entre a questão de pesquisa e aquela que é realmente respondida pelo estudo. Essas diferenças refletem os ajustes que foram feitos para tornar o estudo factível, assim como os erros no delineamento ou na execução do estudo. Há viés quando essas diferenças fazem com que a resposta fornecida pelo estudo seja diferente da resposta correta para a questão de pesquisa. Existem estratégias para minimizar o viés, tanto na fase de delineamento quanto na fase de análise (Tabela 9.2).

- **Fase de delineamento.** Comece escrevendo a questão de pesquisa ao lado do plano de estudo, como ilustrado na Figura 9.1. Então, analise as três questões a seguir, refletindo como elas dizem respeito à questão de pesquisa de interesse:

  1. As **amostras** de sujeitos do estudo (p. ex., casos e controles, ou sujeitos expostos e não expostos) representam adequadamente a(s) população(ões) de interesse?

2. As **medidas das variáveis preditoras** representam adequadamente os preditores de interesse?
3. As **medidas das variáveis de desfecho** representam adequadamente os desfechos de interesse?

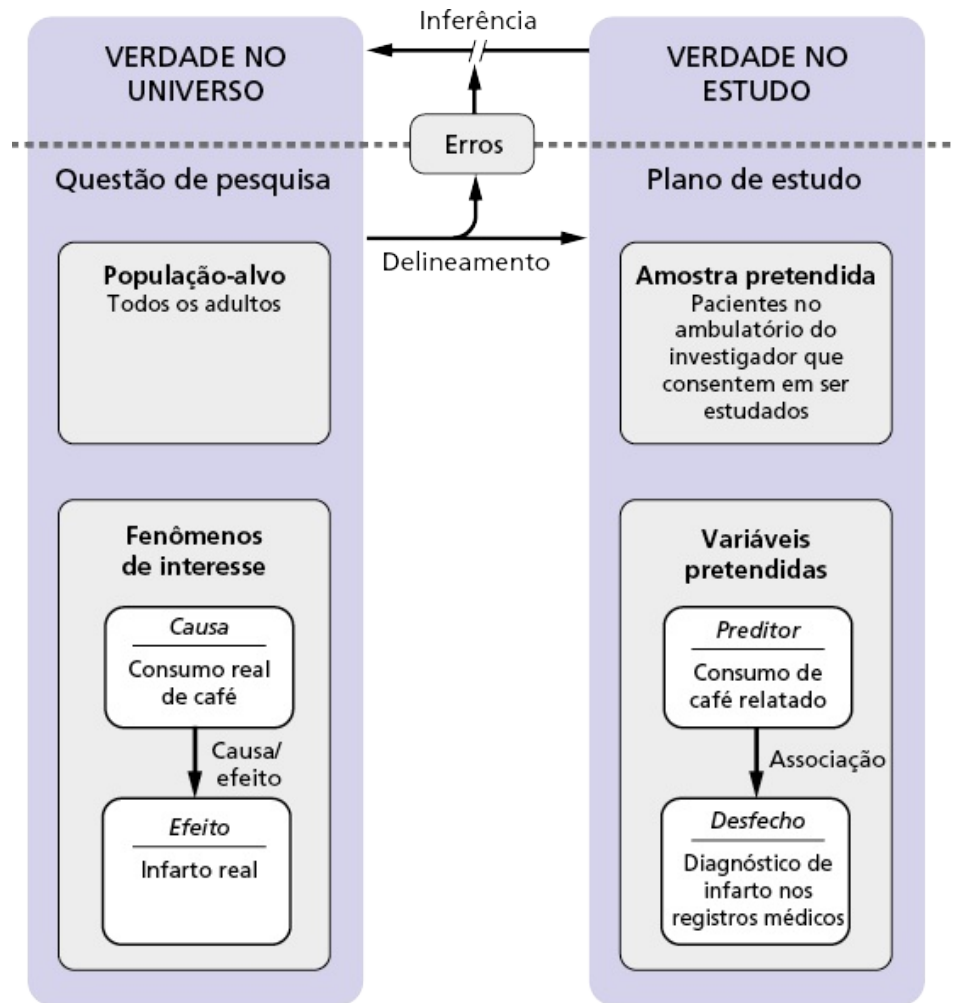



FIGURA 9.1 Como minimizar vieses comparando a questão de pesquisa com o plano de estudo.

Para cada resposta "não" ou "talvez não", é necessário avaliar se o viés é restrito a um dos grupos ou se ele se aplica de forma semelhante a ambos (p. ex., casos e controles ou sujeitos expostos e não expostos) e se ele provavelmente é de magnitude tal que poderia alterar a resposta à questão de pesquisa.

Para ilustrar essas três questões com o exemplo do café e infarto, imagine um estudo de caso-controle no qual os controles sejam

amostrados a partir de pacientes hospitalizados por doenças que não sejam o infarto do miocárdio. Se muitos controles apresentarem doenças crônicas que os levaram a diminuir o consumo de café, a amostra de controles não seria representativa da população que originou os casos. Estariam faltando sujeitos que consomem café. Além disso, se o espasmo esofágico, que pode ser exacerbado pelo café, for diagnosticado erroneamente como infarto, seria observada uma associação espúria entre café e infarto, uma vez que o desfecho medido (diagnóstico de infarto) não representaria corretamente o desfecho de interesse (infarto real).

O passo seguinte é pensar possíveis estratégias para prevenir cada um dos potenciais vieses detectados, como selecionar mais de um grupo-controle em um estudo de caso-controle (Capítulo 8) ou as estratégias para reduzir o viés de aferição descritas no Capítulo 4. Em cada situação, é necessário avaliar a probabilidade de viés e se ele poderia ser facilmente prevenido com mudanças no plano de estudo. Se o viés for facilmente prevenível, revisa-se o plano de estudo e repetem-se as três questões mencionadas. Se o viés não for facilmente prevenível, é hora de se questionar se ainda vale a pena fazer o estudo, avaliando a probabilidade do potencial viés e o quanto ele poderia distorcer a associação que você está tentando estimar.

Pode ser impossível ou muito dispendioso evitar alguns dos potenciais vieses. Além disso, muitas vezes o investigador fica em dúvida sobre até que ponto esses vieses constituem um problema. Nesses casos, deve-se considerar coletar dados adicionais que permitirão avaliar o quanto esses vieses poderão afetar os resultados. Por exemplo, o investigador pode estar preocupado que os casos em um estudo sobre câncer de pâncreas poderiam hiper-relatar exposições recentes a substâncias tóxicas, possivelmente porque estão buscando desesperadamente por uma explicação de por que desenvolveram essa doença. Para lidar com essa possibilidade, o investigador poderia também incluir perguntas sobre exposições (como consumo de café) que estudos prévios já demonstraram não ter efeito sobre o risco de desenvolver esse tipo de câncer. Se o investigador estiver preocupado que um questionário poderia não capturar acuradamente o consumo de café (p. ex., devido a questões mal formuladas), ele pode designar um

entrevistador para examinar cegamente um subconjunto dos casos e controles para determinar se suas respostas concordam com as do questionário. De forma semelhante, se o pesquisador estiver preocupado com a possibilidade de que, em vez de causar o infarto do miocárdio, o café aumente a sobrevivência em pacientes que infartaram (o que poderia as pessoas que tomam café sejam hiper-representadas em uma amostra de pessoas que sobreviveram ao infarto), ele poderia identificar pacientes que infartaram e que morreram e entrevistar seus cônjuges sobre o consumo de café do parceiro.

- **Fase de análise.** Após os dados já terem sido coletados, o objetivo muda de minimizar vieses para analisar o quanto eles poderão afetar os resultados. O primeiro passo é analisar os dados que foram coletados com esse objetivo. Por exemplo, se um investigador antecipa potencial dificuldade de se lembrar se a pessoa tomou ou não café, ele poderia incluir questões sobre a certeza que os casos e os controles têm sobre suas respostas. A associação entre consumo de café e infarto poderia então ser examinada após estratificar de acordo com a certeza sobre o consumo de café, para ver se a associação é mais forte naquelas pessoas que têm maior certeza sobre sua história de exposição.

O investigador também pode examinar os resultados de outros estudos. Se as conclusões forem consistentes, é menos provável que a associação seja resultado de viés. Isso vale especialmente quando os outros estudos usaram delineamentos diferentes e, portanto, provavelmente não compartilharam das mesmas fontes de viés. Entretanto, em muitas situações, os potenciais vieses acabam não sendo um problema significativo. A decisão de como empreender a busca por informações adicionais e como discutir essas questões na redação dos achados do estudo depende do julgamento do investigador. Nesses casos, vale a pena buscar o aconselhamento de colaboradores.

ASSOCIAÇÕES REAIS ALÉM DAS DE CAUSA-EFEITO

Além do acaso e do viés, devem-se considerar os dois tipos de associações que são reais, mas não representam causa-efeito (Tabela 9.3).

TABELA 9.3 Fortalecendo a inferência de que uma associação representa causa e efeito: descartando outras associações reais

TIPO DE ASSOCIAÇÃO REAL	FASE DE DELINEAMENTO (COMO PREVENIR A EXPLICAÇÃO RIVAL)	FASE DE ANÁLISE (COMO AVALIAR A EXPLICAÇÃO RIVAL)
Efeito-causa (o desfecho é, na verdade, causa do preditor)	Realizar um estudo longitudinal para averiguar qual veio primeiro Obter dados sobre a sequência histórica das variáveis (Solução definitiva: realizar um ensaio clínico randomizado)	Avaliar a plausibilidade biológica Comparar a força da associação imediatamente após a exposição ao preditor com a força da associação em algum momento mais tarde Avaliar os achados de outros estudos com delineamentos diferentes
Confundimento (outra variável está associada ao preditor e é causa do desfecho)	Veja Tabela 9.4	Veja Tabela 9.5

Efeito-causa

Uma possibilidade é a carroça ter sido colocada na frente dos bois – o desfecho ser causa do preditor. A relação efeito-causa é um problema em estudos transversais e de caso-controle. Um estilo de vida sedentário causa obesidade, ou o contrário? A relação efeito-causa também pode ser um problema em estudos cruzados de casos. Por exemplo, no estudo sobre o uso de telefones celulares e acidentes automobilísticos descrito no Capítulo 8 (1), o acidente automobilístico poderia ter levado o motorista a fazer uma ligação telefônica relatando o acidente, em vez de o acidente ter sido causado por um motorista desatento. Para abordar essa possibilidade, os investigadores perguntaram aos motoristas sobre o uso do telefone celular antes e depois do acidente e verificaram as respostas usando os registros das ligações telefônicas.

A relação efeito-causa não costuma ser um problema em estudos de coorte sobre a causalidade de doenças, uma vez que as aferições dos fatores de risco podem ser feitas em sujeitos que ainda não desenvolveram a doença em questão. No entanto, mesmo em estudos de coorte, pode haver uma relação efeito-causa se a doença tiver um longo período de latência e os sujeitos com doença subclínica não puderem ser identificados na linha de base. Por exemplo, o diabetes tipo 2 está associado a um risco aumentado de câncer do pâncreas. Parte dessa associação pode muito bem ser de efeito-causa, visto que o câncer do

pâncreas pode afetar as ilhotas pancreáticas que secretam insulina, causando o diabetes. Consistente com a hipótese de causa-efeito, o risco de câncer de pâncreas é mais elevado logo após o diagnóstico do diabetes (2). A associação diminui com o tempo de duração do diabetes (2), mas parte da associação persiste mesmo quatro anos ou mais após o surgimento do diabetes (2-4), sugerindo que pelo menos parte dessa associação possa ser de causa-efeito.

Esse exemplo ilustra uma estratégia geral para se descartar uma relação efeito-causa: avaliar se a associação diminui quando aumenta a duração do intervalo de tempo entre a causa presumida e seu efeito. Uma segunda estratégia é avaliar a plausibilidade biológica da relação efeito-causa quando comparada à relação causa-efeito. Nesse exemplo, a relação efeito-causa era plausível porque o câncer de pâncreas poderia afetar a estrutura do pâncreas, mas a observação de que ter diabetes por mais de 10 anos está associado com um maior risco de uma série de outros tipos de câncer além do câncer de pâncreas (4) aumenta a plausibilidade biológica do diabetes como causa de câncer de pâncreas, em vez de ser apenas um de seus efeitos.

Confundimento

A outra explicação alternativa na Tabela 9.3 é o confundimento, que ocorre quando um terceiro fator é uma causa real do desfecho e a variável preditora de interesse está associada com esse terceiro fator, mas não é a sua causa. Por exemplo, se determinados traços de personalidade fazem com que as pessoas tomem mais café e também tenham maior risco de infarto, esses traços de personalidade irão confundir a associação entre consumo de café e infarto. Se essa for a explicação correta, então a associação entre café e infarto não representa uma relação de causa-efeito, embora seja perfeitamente real. Nesse caso, o consumo de café não passa de um mero espectador na associação.

Para ser confundidora, uma variável precisa estar associada à variável preditora de interesse e também ser uma causa do desfecho. O confundimento pode ser ainda mais complicado, podendo, às vezes, um outro fator também estar envolvido. Por exemplo, o ambiente de trabalho pode levar as pessoas a tomarem café e também a fumarem cigarros, sendo que o tabagismo é um fator de risco para infarto. O Apêndice 9A

fornece um exemplo numérico de como diferenças na taxa de tabagismo poderiam levar a uma associação aparente entre café e infarto.

E se tomar café levar a pessoa a fumar e o fumo for a real causa do infarto? Nesse caso, o tabagismo é denominado **mediador** da associação (causal) entre tomar café e infarto, e não confundidor. Em geral, é melhor evitar controlar para fatores que se encontram na rota causal entre um preditor e um desfecho.

Excluindo o viés, o confundimento muitas vezes é a única explicação alternativa provável à relação de causa-efeito, portanto, é a que se torna mais necessário descartar. É também a mais desafiadora; por esse motivo, estratégias de como lidar com confundidores serão o enfoque do restante deste capítulo. Vale a pena ressaltar, entretanto, que todas essas estratégias envolvem julgamentos, e que não há grau de sofisticação epidemiológica ou estatística que substitua uma compreensão da biologia por trás dos fenômenos observados.

LIDANDO COM CONFUNDIDORES NA FASE DE DELINEAMENTO

A maioria das estratégias para lidar com as variáveis confundidoras exige que elas sejam medidas, e, portanto, uma boa tática é começar listando as variáveis (como idade e sexo) que podem estar associadas à variável preditora e que podem também causar o desfecho. Então, deve-se escolher, entre as estratégias de delineamento e análise, aquelas mais adequadas para controlar a influência dessas potenciais variáveis confundidoras.

As primeiras duas estratégias para a fase de delineamento (Tabela 9.4), **especificação** e **pareamento**, envolvem mudanças no esquema amostral. Os casos e os controles (em um estudo de caso-controle) ou os sujeitos expostos e não expostos (em um estudo de coorte) podem ser selecionados de forma a apresentar valores comparáveis em relação à variável confundidora. Isso elimina a possibilidade de o confundidor ser explicação para uma associação observada entre o preditor e o desfecho. Uma terceira estratégia para a fase de delineamento, que usa **delineamentos oportunísticos**, é apenas aplicável a questões de pesquisa selecionadas, para as quais existem as condições adequadas. No entanto, quando aplicável, esses delineamentos podem se assemelhar aos ensaios clínicos randomizados na sua capacidade de reduzir ou de eliminar o

confundimento, não apenas pelas variáveis aferidas, mas também pelas variáveis não aferidas.

TABELA 9.4 Estratégias de fase de delineamento para lidar com confundidores

ESTRATÉGIA	VANTAGENS	DESVANTAGENS
Especificação	• Fácil compreensão • Direciona a especificação da amostra de sujeitos para a questão de pesquisa	• Limita a capacidade de generalização e de obtenção um tamanho de amostra adequado
Pareamento	• Pode eliminar a influência de importantes confundidores constitucionais, como idade e sexo • Pode eliminar a influência de confundidores difíceis de serem medidos • Pode aumentar o poder estatístico balanceando o número de casos e controles em cada estrato • Pode tornar a amostragem mais conveniente, facilitando a seleção de controles em um estudo de caso-controle	• Pode demandar mais tempo e recursos financeiros, além de ser menos eficiente que aumentar o número de sujeitos (p. ex., o número de controles por acaso) • A decisão de parear deve ser feita no início do estudo e pode afetar de forma irreversível a análise • Requer definição em uma etapa inicial sobre quais variáveis são preditoras e quais são confundidoras • Elimina a opção de estudar as variáveis pareadas como preditores ou como intermediários na rota causal • Requer uma análise para dados pareados • Introduz a possibilidade de hiperpareamento (parear por um fator que não é confundidor, reduzindo, assim, o poder estatístico) • É apenas factível para estudos de caso-controle ou de coortes múltiplas
Delineamentos “oportunísticos”	• Podem aumentar muito a força da inferência causal • Podem ser uma alternativa de menor custo e elegante para o ensaio clínico randomizado	• São possíveis apenas em circunstâncias selecionadas, onde a variável preditora é alocada de forma aleatória ou praticamente aleatória, ou quando existe uma variável instrumental

Especificação

A estratégia mais simples é estabelecer critérios de inclusão que **especifiquem** um valor para uma potencial variável confundidora e, então, excluir todos os indivíduos com valores diferentes. Por exemplo, o investigador que estuda a relação entre café e infarto pode especificar que apenas não fumantes serão incluídos no estudo. Dessa forma, se uma associação for observada entre café e infarto, ela obviamente não se deve ao fumo.

A especificação é uma estratégia eficaz, porém, assim como nos casos de restrições no esquema amostral, apresenta também desvantagens. Em primeiro lugar, mesmo que o café não cause infarto em não fumantes,

pode causar em fumantes. Esse fenômeno – um efeito do café no infarto que é diferente entre fumantes e não fumantes – é denominado **modificação de efeito** (também conhecido como **interação**); ver Apêndice 9A. Portanto, a especificação limita a capacidade de generalização das informações disponíveis em um estudo, comprometendo, nesse exemplo, a capacidade de fazer generalizações para os fumantes. Uma segunda desvantagem é que, se o fumo for bastante prevalente entre os pacientes disponíveis para o estudo, o investigador pode não conseguir recrutar um número adequado de não fumantes. Esses problemas podem se tornar sérios se a especificação for empregada para controlar para um número grande demais de confundidores ou se for empregada de forma muito restritiva. Por exemplo, a capacidade de recrutar amostra com tamanho adequado e de obter ampla generalização dos resultados ficaria comprometida no caso de um estudo restrito a homens de baixa renda, não fumantes e com idades entre 70 e 74 anos.

Pareamento (emparelhamento)

Em um estudo de caso-controle, o **pareamento** pode ser usado para prevenir o confundimento por meio da seleção de casos e controles com os mesmos valores (valores emparelhados) para a(s) variável(is) confundidora(s). Tanto o pareamento quanto a especificação previnem confundimento ao permitirem comparar apenas aqueles casos e controles que compartilham níveis semelhantes do confundidor. O pareamento difere da especificação por preservar a capacidade de generalização, uma vez que os sujeitos de todos os níveis do confundidor são mantidos na amostra em estudo.

O pareamento costuma ser feito individualmente (pareamento **par a par**). Para controlar para tabagismo em um estudo sobre o consumo de café como preditor de infarto, por exemplo, cada caso (um sujeito com infarto) seria individualmente pareado a um ou mais controles com um consumo de cigarros semelhante ao seu (p. ex., 10 a 20 cigarros por dia). O consumo de café do mesmo caso seria, então, comparado com o consumo de café do(s) controle(s) a ele pareado(s).

Outra forma de pareamento é fazê-lo por grupos (**pareamento por frequência**). Para cada nível de fumo, contabilizam-se os casos, então,

seleciona-se um número apropriado de controles com o mesmo nível de fumo. Por exemplo, se o plano do estudo exige dois controles por caso e há 20 casos que fumam 10 a 20 cigarros por dia, os investigadores selecionam 40 controles fumantes desse nível de fumo, pareados como grupo aos 20 casos.

O pareamento é mais comumente usado em **estudos de caso-controle**, mas também pode ser usado em **delineamentos de coortes múltiplas**. Por exemplo, para investigar os efeitos da participação na Guerra do Golfo, entre 1990 e 1991, sobre a fertilidade em militares do sexo masculino, Maconochie e colaboradores (5) compararam homens enviados à região do Golfo durante a guerra com homens que não participaram da guerra, mas que foram pareados por frequência de acordo com o tipo de serviço, a idade, o preparo físico para participar na guerra, e assim por diante. Eles encontraram um risco ligeiramente maior de a pessoa relatar infertilidade (RC de ~ 1,5) e um tempo maior para a concepção entre aqueles que participaram da guerra do Golfo.

Vantagens do pareamento (Tabela 9.4)

- O pareamento é uma forma eficaz de **prevenir o confundimento por fatores constitucionais,** como idade, sexo e raça, que são fortes determinantes do desfecho, não são suscetíveis à intervenção e provavelmente não são intermediários em uma rota causal.
- O pareamento pode ser usado para controlar **confundidores que não podem ser medidos** ou controlados por outra forma. Por exemplo, o pareamento de irmãos (ou, ainda melhor, de gêmeos) pode ajudar a controlar uma série de fatores genéticos e familiares impossíveis de serem medidos. O pareamento por centro clínico, em um estudo multicêntrico, pode controlar para diferenças não especificadas entre as populações ou equipe de pesquisa em centros geograficamente dispersos.
- O pareamento pode aumentar a **precisão** das comparações entre os grupos (e, assim, o poder que o estudo tem para encontrar uma associação real), pois busca um equilíbrio entre o número de casos e o número de controles em cada nível do confundidor. Isso pode ser

importante quando o número disponível de casos é limitado ou quando o custo de estudar os sujeitos é elevado. No entanto, o efeito do pareamento na precisão é modesto e nem sempre favorável (veja mais adiante, no tópico "Desvantagens do pareamento", as informações sobre a possibilidade de hiperpareamento). Em geral, ao se decidir entre parear ou não, o interesse em aumentar a precisão é menos importante do que a necessidade de controlar o confundimento.

- Por fim, o principal uso do pareamento é aumentar a **conveniência**, reduzindo o que seria, de outra forma, um número inviável de potenciais controles. Por exemplo, em um estudo sobre o uso da maconha como fator de risco para neoplasias de testículo de células germinativas, os investigadores pediram aos casos (homens com tumores de testículo) que sugerissem amigos com idade semelhante sem neoplasias para servirem como grupo controle (6). Essa conveniência, no entanto, traz o risco de favorecer o "hiperpareamento".

Desvantagens do pareamento (Tabela 9.4).

- O pareamento demanda **mais tempo e recursos financeiros** para identificar um par para cada sujeito do estudo. Em estudos de caso-controle, por exemplo, à medida que aumenta o número de critérios de pareamento, aumenta também o *pool* de controles necessários para se obter pares para cada caso. O possível aumento em poder estatístico decorrente do pareamento deve ser pesado contra o aumento em poder estatístico que poderia ser obtido arrolando um número maior de casos.
- Quando o pareamento é usado como uma estratégia de amostragem, a decisão de parear deve ser feita no início do estudo, sendo, portanto, **irreversível**. Isso impossibilita uma análise mais detalhada do efeito das variáveis pareadas no desfecho. Pode também resultar em erros sérios se a variável pareada não for constitucional, como idade ou sexo, e sim intermediária na rota causal entre o preditor e o desfecho. Por exemplo, se um investigador que pretendesse investigar os efeitos do consumo de álcool no risco de infarto decidisse parear por níveis de HDL, perderia a parte dos efeitos benéficos do álcool que é mediada por um aumento no HDL. Embora o mesmo erro possa ocorrer nas

estratégias analíticas, o pareamento introduz o erro no estudo de uma forma que não pode mais ser desfeita; nas estratégias analíticas, o erro pode ser evitado alterando a análise.

- A análise correta dos dados pareados exige **técnicas analíticas especiais** (**análises pareadas**) que comparam cada sujeito apenas com seu(s) par(es), e não com outros sujeitos com níveis diferentes de confundidores. Isso significa que os casos para os quais não é possível encontrar um par não podem ser incluídos. No estudo sobre uso de maconha e tumores de células germinativas, 39 dos 187 casos não forneceram um amigo como controle (6). Os autores tiveram de excluir esses 39 casos da análise pareada. O uso de técnicas de análise estatística não pareadas em dados pareados pode produzir resultados incorretos (geralmente com tendência para o não efeito), pois o pressuposto de que os grupos são amostrados de forma independente é violado.
- Uma última desvantagem do pareamento é a possibilidade de **hiperpareamento**, que ocorre quando a variável pareada está associada com a variável preditora mas se mostra como não confundidora por não estar associada ao desfecho. O hiperpareamento pode reduzir o poder estatístico de um estudo de caso-controle, uma vez que a análise pareada descarta os conjuntos de caso-controle pareados com o mesmo nível de exposição (Apêndice 8A.3). No estudo sobre uso de maconha e tumores de células germinativas, por exemplo, o uso amigos controles pode ter reduzido o poder ao aumentar a concordância nas exposições entre casos e seus controles pareados. Os amigos podem tender a ter um padrão semelhante de uso de maconha.

Estudos oportunísticos

Em algumas ocasiões, é possível controlar para variáveis confundidoras na fase de delineamento, mesmo sem medir essas variáveis. Isso é denominado delineamento “oportunístico”, uma vez que ele se aproveita de oportunidades pouco comuns para controlar o confundimento. Um exemplo útil quando se estuda os efeitos imediatos de exposições de curta duração é o estudo **cruzado de casos** (Capítulo 8) – todas as potenciais variáveis confundidoras que são constantes ao longo do tempo (p. ex.,

sexo, raça, classe social, fatores genéticos) são controladas, pois cada sujeito é comparado apenas consigo mesmo em um diferente período no tempo.

Outro delineamento oportunístico envolve um **experimento natural** no qual os sujeitos são expostos ou não expostos a um determinado fator de risco por meio de um processo que, na prática, atua aleatoriamente (7). Por exemplo, Lofgren e colaboradores (8) estudaram os efeitos da descontinuidade dos cuidados hospitalares, aproveitando-se do fato de que os pacientes admitidos após as 17h em sua instituição eram alocados a residentes do último ano que davam prosseguimento aos cuidados desses pacientes ou os transferiam para outra equipe na manhã seguinte. Os autores mostraram que os pacientes cujo atendimento foi transferido tinham 38% mais solicitações de testes laboratoriais ($P = 0{,}01$) e tinham uma mediana de tempo de permanência dois dias maior ($P = 0{,}06$) do que aqueles mantidos na mesma equipe. Da mesma forma, Bell e Redelmeier (9) estudaram os efeitos de características das equipes de enfermagem comparando os desfechos de pacientes com diagnósticos selecionados admitidos em finais de semana com aqueles de pacientes admitidos em dias úteis. Eles encontraram uma maior mortalidade por todas as condições clínicas que eles imaginaram que seriam afetadas pelo menor número de profissionais prestando atendimento nos finais de semana, mas não houve aumento de mortalidade nos pacientes hospitalizados devido a outras condições.

Na medida em que diferenças genéticas na suscetibilidade a uma exposição são elucidadas, uma estratégia denominada **randomização mendeliana** (10) torna-se uma opção. Essa estratégia funciona porque, para polimorfismos genéticos comuns, o alelo que uma pessoa recebe é determinado aleatoriamente dentro de cada família e não está associado à maioria das variáveis confundidoras.

Por exemplo, alguns agricultores que aplicam inseticidas em ovelhas (para eliminar carrapatos, piolhos, etc.) apresentam queixas de saúde, como cefaleia e fadiga, que poderiam estar associadas a essa exposição ocupacional a inseticidas. Investigadores (11) se aproveitaram de um polimorfismo do gene paraoxonase-1, que produz enzimas com capacidades diferentes para hidrolisar o inseticida organofosforado usado em ovelhas (diazinonoxon). Eles descobriram que os agricultores

expostos com queixas relacionadas à sua saúde tinham maior probabilidade de terem alelos associados a uma redução da atividade da paraoxonase-1 do que os agricultores igualmente expostos, porém assintomáticos. Esse achado forneceu evidências fortes para uma relação causal entre exposição a inseticidas usados em ovelhas e problemas de saúde.

Os experimentos naturais e a randomização mendeliana são exemplos de uma abordagem mais geral para melhorar a inferência causal em estudos observacionais, que é o uso de **variáveis instrumentais**. Essas são variáveis associadas ao preditor de interesse, porém não independentemente associadas ao desfecho. O fato de uma pessoa ser admitida em um hospital em um final de semana, por exemplo, está associado ao número de funcionários em serviço, porém não está de outra forma associado com a mortalidade (para os diagnósticos estudados). Portanto, a admissão em um final de semana pode ser considerada uma variável instrumental. Da mesma forma, a atividade da enzima paraoxonase-1 está associada a uma possível toxicidade de inseticidas usados em ovelhas, mas não de outra forma associada a consequências adversas à saúde. Outros exemplos de variáveis instrumentais são o uso de números de sorteio da convocação para serviço militar obrigatório para investigar os efeitos tardios do serviço militar sobre a mortalidade durante a guerra do Vietnã (12) e se a sobrevivência a longo prazo para o estágio inicial do câncer renal depende da distância que um indivíduo mora de um urologista que realiza nefrectomias parciais *versus* um urologista que faz apenas nefrectomias radicais (13).

LIDANDO COM CONFUNDIDORES NA FASE DE ANÁLISE

As estratégias usadas na fase de delineamento – especificação e pareamento – exigem que se decida no início do estudo quais variáveis são confundidoras, não sendo possível estimar posteriormente os efeitos dessas variáveis sobre um desfecho. Por outro lado, as estratégias usadas na fase de análise mantêm as opções em aberto, sendo possível mudar de ideia sobre para quais variáveis se deve controlar no momento da análise.

Pode haver diversas variáveis preditoras, cada uma agindo como confundidora das demais. Por exemplo, embora o consumo de café, o

fumo, o sexo masculino e o tipo de personalidade estejam associados ao infarto, também se associam entre si. Pode-se estipular como meta determinar quais dessas variáveis preditoras estão independentemente associadas ao infarto e quais estão associadas ao infarto apenas por estarem associadas aos outros fatores de risco (causais). Nesta seção, discutiremos os métodos analíticos para avaliar a contribuição **independente** das variáveis preditoras em estudos observacionais. Esses métodos são resumidos na Tabela 9.5.[1]

TABELA 9.5 Estratégias da fase de análise para lidar com confundidores

ESTRATÉGIA	VANTAGENS	DESVANTAGENS
Estratificação	• Fácil compreensão • Flexível e reversível; a escolha sobre a partir de quais variáveis estratificar pode ser feita após a coleta dos dados	• O número de estratos é limitado pelo tamanho de amostra necessário para cada estrato: • Poucas covariáveis podem ser consideradas simultaneamente • O número limitado de estratos por covariável leva a um controle parcial do confundimento • É necessário que as covariáveis relevantes já tenham sido medidas

Ajuste estatístico	• É possível controlar múltiplos confundidores simultaneamente
	• A informação contida nas variáveis contínuas pode ser usada em sua completude
	• Flexível e reversível
		• O modelo pode não ter ajuste adequado:
		  • Controle incompleto do confundimento (caso o modelo não se ajuste à relação confundidor-desfecho)
		  • Estimativas errôneas da magnitude do efeito (caso o modelo não se ajuste à relação preditor-desfecho)
		• Os resultados podem ser de difícil compreensão (muitas pessoas não compreendem prontamente o significado de coeficiente de regressão.)
		• É necessário que as covariáveis relevantes já tenham sido medidas

Escores de propensão	• É possível controlar múltiplos confundidores simultaneamente • A informação contida nas variáveis contínuas pode ser usada em sua completude • Aumentam o poder estatístico para controlar para o confundimento quando um número maior de pessoas recebem o tratamento do que o número que desenvolve o desfecho • Se uma análise estratificada ou pareada for usada, não é necessário observar pressupostos do modelo • Flexíveis e reversíveis • Falta de sobreposição dos escores de	• Os resultados podem ser de difícil compreensão • É necessário que as covariáveis relevantes já tenham sido medidas • Só podem ser usados quando os escores de propensão dos sujeitos expostos e não expostos se sobrepoem, o que reduz o tamanho de amostra

propensão pode revelar subgrupos nos
quais é difícil ou impossível controlar para
o confundimento

Estratificação

Assim como na especificação e no pareamento, a **estratificação** também assegura que sejam comparados somente casos e controles (ou expostos e não expostos) com níveis semelhantes de um potencial confundidor. A estratificação segrega os sujeitos em estratos **(subgrupos)** de acordo com o nível do potencial confundidor e, então, examina a relação entre o preditor e o desfecho separadamente em cada estrato. A estratificação é ilustrada no Apêndice 9A. Analisando-se fumantes e não fumantes separadamente ("estratificação pelo fumo"), é possível eliminar os efeitos confundidores do fumo.

O Apêndice 9A também ilustra a **modificação de efeito**, situação na qual a estratificação revela que a associação entre a variável preditora e a de desfecho se altera de acordo com o nível de um terceiro fator (ou seja, é modificada por ele). A modificação de efeito introduz uma complexidade adicional, pois passa a não ser mais possível usar uma única medida de associação para sumarizar a relação entre a variável preditora e a de desfecho. Devido ao acaso, raramente as estimativas de associação em diferentes estratos serão idênticas; por esse motivo, os achados somente irão sugerir modificação de efeito se as estimativas variarem muito. Não é comum haver modificação de efeito clinicamente significativa e, antes de concluir que ela está presente, deve-se avaliar sua significância estatística. Além disso, é importante avaliar se esse achado pode ser replicado em outra população, especialmente quando muitos grupos são examinados, o que aumenta a probabilidade de que pelo menos um deles seja estatisticamente significativo devido ao acaso. A plausibilidade biológica, ou a falta dela, também pode contribuir para a interpretação. A questão da modificação de efeito também surge em análises de subgrupos em ensaios clínicos (Capítulo 11) e em metanálises quando se considera a homogeneidade (similaridade) de estudos (Capítulo 13).

A estratificação tem a vantagem da **flexibilidade**, pois, ao fazer diversas análises estratificadas, o investigador pode decidir que variáveis parecem ser confundidoras e ignorar as demais. Isso pode ser feito

combinando o conhecimento sobre as prováveis direções das relações causais com análises para averiguar se os resultados das análises estratificadas diferem substancialmente dos resultados de análises não estratificadas (ver o Apêndice 9A). A estratificação também tem a vantagem de ser reversível. Não é necessário tomar decisões no início do estudo que podem, mais tarde, causar arrependimento.

A principal desvantagem da análise estratificada é o número limitado de variáveis que podem ser controladas simultaneamente. Por exemplo, os possíveis confundidores no estudo do café e infarto podem incluir idade, tipo de personalidade, pressão arterial sistólica, colesterol sérico e consumo de cigarros. Para estratificar de acordo com essas cinco variáveis, com apenas três estratos para cada uma delas, seriam necessários $3^5 = 243$ estratos. Com esse número de estratos, alguns deles não teriam nenhum caso ou nenhum controle, impossibilitando seu uso.

Para manter um número suficiente de sujeitos em cada estrato, é comum dividir uma variável em estratos mais amplos. No entanto, quando os estratos são muito amplos, o confundimento pode não ser controlado de forma adequada. Por exemplo, se o estudo mencionado estratificasse a amostra pela idade, usando apenas dois estratos (p. ex., idade < 50 e idade ≥ 50 anos), ainda poderia haver confundimento residual, caso os sujeitos com maior consumo de café em cada estrato de idade fossem aqueles com maior idade e, dessa forma, tivessem um maior risco de infarto.

Ajuste

Há várias técnicas estatísticas disponíveis para **ajustar** para confundidores. Essas técnicas **modelam** a natureza das associações entre as variáveis para poder isolar os efeitos das variáveis preditoras daqueles de variáveis confundidoras. Por exemplo, um estudo que investiga o efeito da ingestão de chumbo sobre o quociente de inteligência (QI) de crianças poderia examinar o nível de escolaridade dos pais como potencial confundidor. O ajuste estatístico poderia modelar a relação entre a escolaridade dos pais e o QI da criança supondo uma linha reta, na qual cada ano de escolaridade dos pais está associado a um aumento fixo no QI da criança. Os QIs das crianças com níveis diferentes de chumbo poderiam, então, ser ajustados para eliminar o efeito da escolaridade dos

pais usando-se a estratégia analítica descrita no Apêndice 9B. Com frequência, um investigador quer ajustar simultaneamente para vários potenciais confundidores – como idade, sexo, raça e escolaridade. Para isso, é necessário utilizar técnicas para ajuste multivariado, como regressão linear multivariável, regressão logística ou análise de azares proporcionais de Cox. Essas técnicas têm outra vantagem: permitem utilizar toda a informação contida nas variáveis contínuas. É fácil, por exemplo, ajustar para o nível de escolaridade do pai ou da mãe em intervalos de um ano, em vez de estratificar a amostra em apenas poucas categorias de grau de escolaridade. Além disso, **termos de interação** podem ser usados para modelar a modificação de efeito entre as variáveis.

No entanto, o ajuste multivariado apresenta desvantagens. A mais importante é que o modelo pode não funcionar. Os pacotes de programas estatísticos tornaram esses modelos tão acessíveis que o investigador pode não parar para refletir se o seu uso é adequado para as variáveis preditora e de desfecho em questão.[2] No exemplo do Apêndice 9B, o investigador deve examinar os dados para avaliar se a relação entre o grau de escolaridade dos pais e o QI da criança é realmente linear. Se o padrão for muito diferente (p. ex., a inclinação da reta se tornando mais acentuada com o aumento da escolaridade), as tentativas de ajustar o QI para escolaridade dos pais usando um modelo linear serão imperfeitas e a estimativa do efeito independente do chumbo será incorreta.

Em segundo lugar, as estatísticas são frequentemente de difícil compreensão. Isso ocorre especialmente quando são utilizadas transformações de variáveis (p. ex., escolaridade dos pais ao quadrado) ou termos de interação. Os investigadores devem passar o tempo que for necessário com um estatístico (ou realizar os cursos necessários) para assegurar que conseguirão explicar o significado dos coeficientes ou de outras estatísticas altamente derivadas que eles planejam relatar. Como medida de precaução, é uma boa ideia sempre começar com análises simples, estratificadas, e buscar auxílio para compreender o que está ocorrendo se as análises mais complicadas produzirem resultados substancialmente diferentes.

Escores de propensão

Os **escores de propensão** podem ser particularmente úteis em estudos

observacionais sobre a eficácia de tratamentos, para controlar para o **confundimento por indicação** – isto é, para o fato de que os pacientes para os quais um tratamento é indicado (e prescrito) costumam ser de maior risco ou de outra forma diferentes do que aqueles que não recebem o tratamento. É importante lembrar que, para ser confundidora, uma variável deve estar associada tanto ao preditor quanto ao desfecho. Em vez de ajustar para todos os fatores que predizem o desfecho, o uso do escore de propensão envolve a criação de um modelo multivariado para predizer quem receberá o tratamento. A cada sujeito, então, é atribuída uma probabilidade de receber o tratamento – um “escore de propensão”. Esse escore pode ser usado como única variável confundidora em uma análise estratificada ou multivariada.

Como outra opção, os sujeitos que receberam e os que não receberam o tratamento poderiam ser pareados por escore de propensão, comparando-se os desfechos entre os pares. Ao contrário do uso do pareamento como estratégia da fase de delineamento, o pareamento por propensão se assemelha a outras estratégias de fase de análise por ser reversível. Entretanto, as análises pareadas por propensão não são adequadas para sujeitos que não podem ser pareados porque seus escores de propensão são próximos de 0 ou de 1. Embora isso reduza o tamanho de amostra, pode ser vantajoso, pois nesses sujeitos que não podem ser pareados a análise por escore de propensão terá identificado uma falta de comparabilidade entre grupos e incapacidade de controlar para confundimento que não teria sido aparente com outros métodos de análise multivariável.

O uso de escores de propensão apresenta diversas vantagens. O número de potenciais variáveis confundidoras que podem ser modeladas como preditoras de uma intervenção é geralmente maior do que o número de variáveis que podem ser modeladas como preditoras de um desfecho, uma vez que o número de indivíduos tratados é, na maioria das vezes, muito superior ao número que desenvolve o desfecho (2.310 comparado a 276 no Exemplo 9.1). Outro motivo pelo qual um número maior de confundidores pode ser incluído é que não há risco de “hiperajustar” o modelo de propensão – podem-se incluir termos de interação, termos quadráticos e múltiplas variáveis indicadoras (15). Por fim, os investigadores geralmente têm mais confiança para identificar os

determinantes do tratamento do que os determinantes do desfecho, pois os médicos tomam decisões terapêuticas com base em um número limitado de características dos pacientes.

EXEMPLO 9.1 Análise de propensão

Gum e colaboradores (14) estudaram prospectivamente 6.174 adultos submetidos à ecocardiografia de esforço, 2.310 dos quais (37%) estavam tomando aspirina e 276 dos quais morreram no período de seguimento de 3,1 ano. Nas análises não ajustadas, o uso de aspirina não esteve associado à mortalidade (4,5% em ambos os grupos). No entanto, quando os 1.351 pacientes que receberam aspirina foram pareados com os 1.351 pacientes com o mesmo escore de propensão, mas que não receberam aspirina, a mortalidade foi 47% inferior no grupo tratado ($P = 0{,}002$).

Assim como ocorre com outras técnicas multivariadas, para se usar escores de propensão, é necessário que as potenciais variáveis confundidoras sejam identificadas e medidas. Uma **limitação** dessa técnica é que ela não fornece informações sobre a relação entre cada variável confundidora isoladamente e o desfecho – o único resultado fornecido é para a variável preditora que foi modelada (geralmente um tratamento). No entanto, como essa é uma estratégia de fase de análise, ela não impede que sejam feitas também análises multivariadas mais tradicionais; na prática, ambos os tipos de análise são em geral empregados juntos.

OUTRAS ARMADILHAS AO QUANTIFICAR OS EFEITOS CAUSAIS

Condicionamento em um efeito comum

O viés devido ao **condicionamento em um efeito comum** é um tópico traiçoeiro e frequentemente omitido em livros-texto introdutórios, uma vez que a maioria das explicações usam diagramas e notação abstratos. Abordaremos este tópico de uma forma diferente, começando com exemplos de como esse viés poderia ocorrer e então tentando explicar o significado de seu nome.

Considere um estudo sobre pessoas que perderam pelo menos 7 kg no último ano. Um investigador descobre que os sujeitos que estavam fazendo dieta tinham um risco menor de câncer do que aqueles que não estavam fazendo dieta. Você acha que fazer dieta preveniu o câncer nesses sujeitos?

Se você parar para pensar, provavelmente responderá que não, uma vez que o câncer também leva à perda de peso. Você pode imaginar que, se alguém perder peso sem causa aparente, é muito mais provável que isso signifique um câncer do que se alguém perder peso durante a dieta. Entre pessoas que perderam peso, se a perda de peso não foi causada pela dieta, é mais provável que tenha sido causada por algo mais grave. Os investigadores criaram uma associação inversa entre fazer dieta e câncer por meio do condicionamento (restrição da atenção) em um efeito comum (perda de peso, que é causada tanto pela dieta quanto pelo câncer).

A seguir descreveremos outro exemplo. Entre bebês com baixo peso ao nascer, aqueles cujas mães fumaram durante a gestação têm menor mortalidade infantil do que aqueles cujas mães não fumaram (16). Devemos encorajar que mais mães fumem durante a gestação? Certamente não! O motivo dessa observação é que o fumo leva ao baixo peso ao nascer, mas há também outras causas para esse desfecho, especialmente a prematuridade. Portanto, entre bebês com baixo peso ao nascer, se esse baixo peso não foi causado pelo tabagismo, é mais provável que tenha sido causado pela prematuridade. Os investigadores criaram uma associação inversa entre fumo e prematuridade (e seu risco de mortalidade associado) condicionando em (restringindo a atenção para) um efeito comum (baixo peso ao nascer, que é causado tanto pelo fumo quanto pela prematuridade).

Agora a expressão “condicionamento em um efeito comum” faz sentido. **Condicionamento** é um termo epidemiológico que significa examinar associações entre variáveis preditoras e de desfecho “condicionadas em” (isto é, em níveis especificados de) algum atributo. **Efeito comum** se refere a um atributo (como perder peso ou ser um bebê com baixo peso ao nascer) que tem diversas causas. O viés devido ao condicionamento em um efeito comum pode ocorrer se o investigador tratar algo causado pelo fator de risco que está sendo estudado como um critério de inclusão, uma variável de pareamento ou uma possível variável

confundidora.

Subestimativa de efeitos causais

Até agora, nossa ênfase foi em avaliar a probabilidade de explicações alternativas para uma associação, de modo a evitar uma conclusão falsa de que uma associação é real e causal quando ela na verdade não é. No entanto, outro tipo de erro também é possível: o de *subestimar* efeitos causais. O acaso, o viés e o confundimento podem também ser razões para deixar passar ou subestimar uma associação.

Discutimos o **acaso** como motivo para se deixar de encontrar uma associação no Capítulo 5, quando abordamos erros tipo II e a necessidade de assegurar que o tamanho de amostra irá fornecer um **poder estatístico** adequado para encontrar associações reais. Após um estudo ter sido concluído, no entanto, o cálculo do poder estatístico não é mais uma boa forma de quantificar a incerteza provocada pelo erro aleatório. Nesse momento, o poder hipotético de um estudo para detectar um efeito de uma determinada magnitude é menos relevante do que os achados propriamente ditos, expressos como a estimativa observada da associação (p. ex., risco relativo) e seu **intervalo de confiança** de 95% (17).

O **viés** também pode distorcer as estimativas de uma associação em direção à ausência de efeito. No Capítulo 8, o cegamento na avaliação da presença ou da ausência de um fator de risco em casos e controles ocorreu para evitar o **viés de aferição diferencial**, que ocorre, por exemplo, quando há diferenças entre os casos e os controles na forma como as questões são formuladas ou como as respostas são interpretadas, que podem levar os observadores a obter a resposta que desejam. Como os observadores podem desejar resultados em qualquer uma das direções, o viés de aferição diferencial pode enviesar os resultados de modo a hiperestimar ou subestimar efeitos causais. O viés não diferencial, por outro lado, geralmente leva a subestimar as associações.

O **confundimento** também pode levar a uma atenuação de associações reais. Por exemplo, suponha que tomar café realmente proteja contra o infarto, mas seja mais comum em fumantes. Se não houver controle para o tabagismo, os efeitos benéficos do café podem não ser percebidos – pessoas que tomam café podem aparentar ter o mesmo risco de infarto que aquelas que não tomam, quando sua maior prevalência de tabagismo

deveria ter feito com que seu risco fosse maior. Esse tipo de confundimento, no qual os efeitos de um fator benéfico são mascarados por sua associação com uma causa do desfecho, é às vezes denominado **supressão** (18). É um problema comum em estudos observacionais sobre tratamentos, pois os tratamentos costumam ser mais indicados justamente naqueles indivíduos de maior risco para um desfecho ruim. O resultado, como discutido anteriormente, é que um tratamento benéfico pode parecer sem utilidade nenhuma (como foi o caso da aspirina no Exemplo 9.1) ou mesmo aparentar provocar danos, até que o confundimento por indicação tenha sido controlado.

ESCOLHENDO UMA ESTRATÉGIA

Que diretrizes gerais podem ser oferecidas para decidir entre lidar com os confundidores na fase de delineamento ou na fase de análise? É mais adequado usar a **especificação** para controlar o confundimento quando o investigador tem maior interesse em subgrupos específicos da população. A especificação nada mais é que uma forma especial de um processo geral, que é estabelecer critérios de seleção da amostra (Capítulo 3). Entretanto, para estudos nos quais a inferência causal é a meta, há a preocupação adicional de evitar critérios de inclusão que poderiam ser causados por variáveis preditoras que você quer estudar (isto é, condicionamento em um efeito comum).

Uma decisão importante no delineamento do estudo é sobre fazer ou não **pareamento**. O pareamento é mais adequado para fatores constitucionais de valores fixos, como idade, raça e sexo. Pode também ser útil quando o tamanho de amostra é pequeno se comparado ao número de estratos necessários para o controle de confundidores conhecidos, e quando os confundidores são mais facilmente pareados do que medidos. A desvantagem do pareamento é que pode comprometer permanentemente a capacidade do investigador em observar associações reais, por isso, deve-se usá-lo com parcimônia, especialmente para variáveis que podem estar na rota de causalidade. Em muitas situações, as estratégias da fase de análise (estratificação, ajuste e escores de propensão) são igualmente adequadas para controlar o confundimento, tendo a vantagem de ser **reversíveis** – permitem que o investigador

adicione ou retire covariáveis para explorar diferentes modelos causais.

Embora ele não esteja disponível para todas as questões de pesquisa, é sempre interessante considerar a possibilidade de um delineamento oportunístico. Se você não parar para considerar esse tipo de estudo (e perguntar aos seus colaboradores a respeito dele), você poderá perder uma grande oportunidade de fazê-lo.

A decisão final de **estratificar**, **ajustar** ou usar **escores de propensão** não precisa ser feita até ter sido concluída a coleta dos dados; muitas vezes, o investigador irá querer usar todas essas estratégias. No entanto, é importante, durante o delineamento do estudo, refletir sobre que fatores poderão ser usados para o ajuste, de forma a saber que variáveis medir. Além disso, uma vez que diferentes estratégias da fase de análise para controlar o confundimento nem sempre produzem os mesmos resultados, é melhor especificar um plano principal de análise *a priori*. Isso pode ajudar os investigadores a resistir à tentação de selecionar a estratégia que fornece o resultado mais desejável.

Evidências que favorecem a causalidade

As abordagens para fortalecer a inferência causal têm sido, até agora, em grande parte, abordagens de negação, isto é, de se refutar as quatro explicações rivais da Tabela 9.1. Uma estratégia complementar é identificar características das associações que forneçam evidências positivas para a causalidade. Entre elas, as mais importantes são a consistência e a magnitude da associação, a presença de uma relação dose-resposta e a plausibilidade biológica.

Quando os resultados são **consistentes** entre estudos com diferentes delineamentos, é menos provável que o acaso ou vieses sejam a causa de uma associação. No entanto, associações reais que representam efeito-causa ou confundimento também podem ser observadas de forma consistente entre vários estudos. Por exemplo, se os indivíduos que fumam cigarros consomem mais café e têm maior incidência de infarto, os estudos demonstrarão consistentemente uma associação entre consumo de café e infarto.

A **magnitude** da associação também é importante. Em primeiro lugar, uma associação mais forte fornece um valor *P* de maior significância estatística, tornando o acaso uma explicação menos provável. Associações

mais fortes também fornecem evidências positivas de causalidade, pois reduzem a probabilidade de confundimento. As associações induzidas pelo confundimento são indiretas (via confundidor) e, dessa forma, são geralmente mais fracas que as associações diretas de causa-efeito. Isso é ilustrado no Apêndice 9A, onde as associações fortes entre café e fumo (RC = 16) e entre fumo e infarto (RC = 4) produziram uma associação bem mais fraca entre café e infarto (RC = 2,25).

Uma relação de **dose-resposta** também fornece evidências positivas para a causalidade. Um exemplo disso é a associação entre consumo de cigarros e câncer do pulmão. Fumantes moderados têm índices mais elevados de câncer que não fumantes, e indivíduos que fumam muito têm índices ainda mais elevados. Sempre que possível, as variáveis preditoras devem ser medidas em seu contínuo ou em várias categorias, de forma que qualquer relação dose-resposta presente possa ser observada. No entanto, é possível observar uma relação dose-resposta com associações efeito-causa ou na presença de confundimento.

Por fim, é importante também considerar a **plausibilidade biológica**. Pode-se fortalecer as evidências para a causalidade quando é possível propor um mecanismo causal que faça sentido biologicamente. Se as associações não fizerem sentido, a probabilidade de representarem causa-efeito será menor. Por exemplo, no estudo sobre uso de maconha como fator de risco para tumores de células germinativas, o uso de maconha menos de uma vez por dia estava associado com um menor risco do que nenhum uso (6). É difícil explicar esse achado biologicamente.

A plausibilidade biológica não deve ser hiperenfatizada. É possível propor mecanismos para praticamente qualquer associação, e algumas associações inicialmente vistas como pouco plausíveis biologicamente, como a etiologia bacteriana para a úlcera péptica, se mostraram verdadeiras.

RESUMO

1. Ao delinear **estudos observacionais**, deve-se ter em mente a necessidade de interpretar as **associações** encontradas. A inferência de que a associação representa uma relação de **causa-efeito** é fortalecida por estratégias que reduzam a probabilidade das **quatro explicações**

rivais – **acaso, viés**, **efeito-causa** e **confundimento**.

2. O papel do **acaso (erro aleatório)** pode ser minimizado delineando-se um estudo com **tamanho de amostra e precisão adequados** para garantir um baixo índice de **erro tipo I** e de **erro tipo II**. Uma vez terminado o estudo, o efeito do erro aleatório pode ser avaliado a partir da amplitude do **intervalo de confiança de 95%** e da consistência entre os resultados do estudo e as **evidências anteriormente publicadas**.
3. O **viés (erro sistemático)** surge de diferenças entre, por um lado, a população e os fenômenos incluídos na questão de pesquisa e, por outro, os sujeitos e as medidas reais do estudo. Para tentar minimizá-lo no delineamento, é importante **julgar** se essas diferenças produziriam uma resposta errada para a questão de pesquisa.
4. É possível tornar a relação **efeito-causa** menos provável delineando-se um estudo que permita a avaliação da **sequência temporal** e considerando a **plausibilidade biológica**.
5. O **confundimento**, que pode estar presente quando uma terceira variável está associada com o preditor de interesse e é causa do desfecho, torna-se menos provável por meio das seguintes estratégias, a maioria das quais requer que os potenciais confundidores sejam antecipados e medidos:
   a. **Especificação** ou **pareamento**, duas estratégias da **fase de delineamento** que alteram o plano amostral para garantir que sejam comparados apenas grupos com níveis semelhantes do confundidor. Essas estratégias **devem ser usadas com parcimônia**, uma vez que podem limitar de forma **irreversível** a informação disponível no estudo.
   b. Algumas **estratégias de fase de análise** que alcançam o mesmo objetivo e preservam opções para investigar rotas causais são:

      - **Estratificação**, que, além de controlar para o confundimento, pode revelar **modificação de efeito (“interação”)**, ou seja, uma magnitude diferente da associação entre preditor e desfecho em níveis diferentes de uma terceira variável.

      - **Ajuste**, que permite que o impacto de muitas variáveis preditoras

seja controlado simultaneamente.

- **Escores de propensão**, que aumentam o poder estatístico para abordar o **confundimento por indicação** em estudos observacionais sobre a eficácia de um tratamento.

6. Os investigadores devem ficar atentos para a possibilidade de usar delineamentos observacionais **oportunísticos**, incluindo **experimentos naturais**, **randomização mendeliana** e outros delineamentos de **variáveis instrumentais** que permitam inferir sobre a causalidade de forma semelhante à proporcionada por um ensaio clínico randomizado.
7. Os investigadores devem evitar o **condicionamento em um efeito comum** na fase de delineamento, ao não selecionar sujeitos com base em covariáveis que poderiam ser causadas pelo preditor e, na fase de análise, ao não controlar para essas covariáveis.
8. A inferência causal pode ser fortalecida ainda mais por evidências positivas: **consistência e magnitude da associação**, presença de uma relação **dose-resposta** e **plausibilidade biológica**.

APÊNDICE 9A

Exemplo hipotético de confundimento e modificação de efeito

Os números nas tabelas a seguir correspondem aos números de sujeitos nesse estudo de caso-controle hipotético.

PAINEL 1. Se examinarmos o conjunto total dos sujeitos, parece haver uma associação entre consumo de café e infarto do miocárdio:

	Fumantes e não fumantes combinados	
	Com infarto	Sem infarto
Consumidor de café	90	60
Não consumidor de café	60	90

RC para a associação entre infarto do miocárdio e café em fumantes e não fumantes combinados $= \dfrac{90 \times 90}{60 \times 60} = 2{,}25$

PAINEL 2. No entanto, isso pode se dever ao **confundimento**, como mostrado a seguir nas tabelas estratificadas pelo fumo. Estas tabelas mostram que o consumo de café não está associado ao infarto nem em fumantes nem em não fumantes:

	Fumantes	
	Com Infarto	Sem infarto
Consumidor de café	80	40
Não consumidor de café	20	10

	Não fumantes	
	Com Infarto	Sem infarto
Consumidor de café	10	20
Não consumidor de café	40	80

RC para a associação entre infarto e café:

Em fumantes $= \dfrac{80 \times 10}{20 \times 40} = 1$

Em não fumantes $= \dfrac{10 \times 80}{40 \times 20} = 1$

O tabagismo é um confundidor, pois está fortemente associado ao consumo de café (ver a seguir, no painel à esquerda) e com o infarto do

miocárdio (ver a seguir, no painel à direita). Estas tabelas foram obtidas pelo rearranjo dos números no Painel 2.

	Com infarto e sem infarto, combinados	
	Café	Sem café
Fumantes	120	30
Não fumantes	30	120

	Consumidores de café e não consumidores de café, combinados	
	Café	Sem café
Fumantes	100	50
Não fumantes	50	100

RC para o consumo de café associado ao tabagismo

$$= \frac{120 \times 120}{30 \times 30} = 16$$

RC para o infarto associado ao tabagismo

$$= \frac{100 \times 100}{50 \times 50} = 4$$

PAINEL 3. A associação entre consumo de café e infarto no Painel 1 também poderia representar **modificação de efeito**, se a estratificação pelo fumo revelasse que a associação entre consumo de café e infarto difere em fumantes e não fumantes. Na tabela a seguir, a RC de 2,25 para a associação entre consumo de café e infarto em fumantes e não fumantes, combinados, deve-se inteiramente à associação forte em fumantes. Quando uma interação está presente, as RCs em diferentes estratos são diferentes, devendo ser relatadas separadamente:

	Fumantes	
	Com Infarto	Sem infarto
Tomam café	50	15
Não tomam café	10	33

	Não fumantes	
	Com Infarto	Sem infarto
Tomam café	40	45
Não tomam café	50	57

As RCs para infarto associado ao consumo de café são:

$$\text{RC em fumantes} = \frac{50 \times 33}{15 \times 10} = 11 \qquad \text{RC em não fumantes} = \frac{40 \times 57}{45 \times 50} = 1$$

Conclusão: A associação entre consumo de café e infarto no Painel 1 poderia estar mascarando a presença do confundimento pelo fumo, que seria revelado estratificando pela variável fumo (Painel 2). Poderia também estar mascarando a presença de modificação de efeito pelo fumo, que também seria revelada estratificando por essa variável (Painel 3). Além disso, poderia representar uma associação de causa-efeito, que seria

apoiada (embora sem comprovação definitiva) se a estratificação pelo fumo não alterasse a associação entre consumo de café e infarto. Uma última alternativa (mais realista) é essa associação ser resultado de uma mistura dos fatores citadas.

APÊNDICE 9B
Exemplo simplificado de ajuste

Imagine que um estudo encontre dois fortes preditores do QI em crianças: o nível de escolaridade dos pais e o nível de chumbo no sangue da criança. Considere os seguintes dados hipotéticos sobre crianças com níveis normais e elevados de chumbo:

	Média de escolaridade dos pais (em anos)	QI médio da criança
Níveis elevados de chumbo	10	95
Níveis normais de chumbo	12	110

Como o grau de escolaridade dos pais também está associado ao nível sérico de chumbo na criança, é preciso saber se a diferença em QI associada ao nível de chumbo sérico nas crianças é maior do que aquela explicada pela diferença no grau de escolaridade dos pais. Para responder essa pergunta, é preciso saber que diferença em QI se espera que a diferença no grau de escolaridade produza. Isso pode ser feito plotando-se o grau de escolaridade dos pais contra o QI dos filhos com níveis normais de chumbo (Figura 9.2)[3].

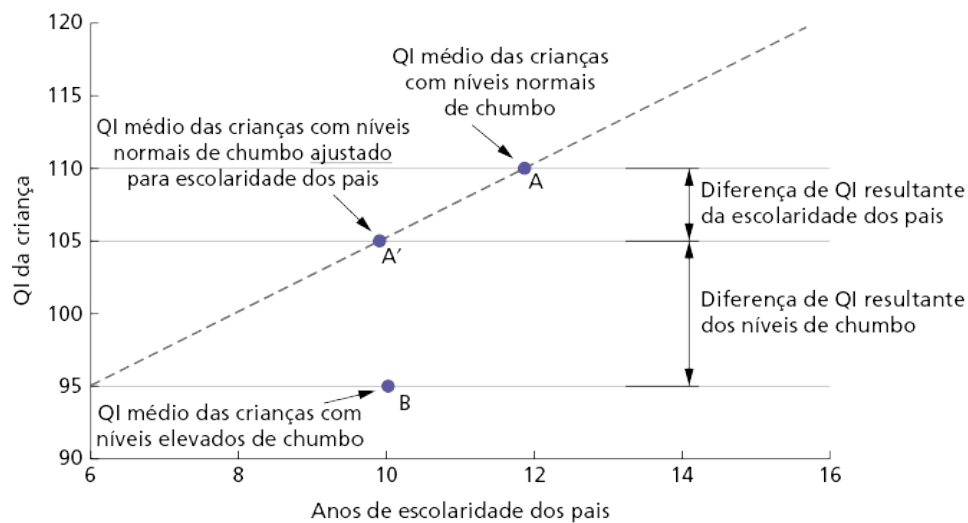



FIGURA 9.2 Gráfico hipotético do QI como função linear (linha pontilhada) da escolaridade dos pais (em anos).

A linha diagonal pontilhada na Figura 9.2 mostra a relação entre o QI da criança e a escolaridade dos pais em crianças com níveis normais de chumbo: o QI da criança aumenta 5 pontos para cada incremento de dois anos na escolaridade dos pais. Dessa forma, pode-se ajustar o QI no grupo com níveis normais de chumbo para controlar a diferença em escolaridade dos pais indo do ponto *A* para o ponto *a*'. (Uma vez que o grupo com níveis normais de chumbo tinha, em média, dois anos a mais de escolaridade dos pais, ajustamos seus QIs para baixo em 5 pontos, tornando-os comparáveis em termos de escolaridade dos pais aos QIs do grupo com níveis elevados de chumbo.) Isso ainda deixa uma diferença de QI de 10 pontos entre os pontos *A* e *B*, sugerindo que o chumbo tem um efeito independente, dessa magnitude, no QI das crianças. Portanto, da diferença de 15 pontos no QI de crianças com níveis baixos e elevados de chumbo, 5 pontos podem ser explicados pela diferença de escolaridade dos pais, e os 10 restantes, pela exposição das crianças ao chumbo.

REFERÊNCIAS

1. McEvoy SP, Stevenson MR, McCartt AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ* 2005;331(7514):428.
2. Magruder JT, Elahi D, Andersen DK. Diabetes and pancreatic cancer: chicken or egg? *Pancreas* 2011;40(3):339–351.
3. Huxley R, Ansary-Moghaddam A, Berrington de Gonzalez A, et al. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. *Br J Cancer* 2005;92(11):2076–2083.
4. Bosetti C, Rosato V, Polesel J, et al. Diabetes mellitus and cancer risk in a network of case-control studies. *Nutr Cancer* 2012;64(5):643–651.
5. Maconochie N, Doyle P, Carson C. Infertility among male UK veterans of the 1990-1 Gulf war: reproductive cohort study. *BMJ* 2004;329(7459):196–201.
6. Trabert B, Sigurdson AJ, Sweeney AM, et al. Marijuana use and testicular germ cell tumors. *Cancer* 2011;117(4):848–853.
7. Newman TB, Kohn M. *Evidence-based diagnosis*. New York: Cambridge University Press, 2009. Chapter 10.
8. Lofgren RP, Gottlieb D, Williams RA, et al. Post-call transfer of resident responsibility: its effect on patient care [see comments]. *J Gen Intern Med* 1990;5(6):501–505.
9. Bell CM, Redelmeier DA. Mortality among patients admitted to hospitals on weekends as compared with weekdays. *N Engl J Med* 2001;345(9):663–668.
10. Davey Smith G, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology

contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32(1):1–22.

11. Cherry N, Mackness M, Durrington P, et al. Paraoxonase (PON1) polymorphisms in farmers attributing ill health to sheep dip. *Lancet* 2002;359(9308):763–764.
12. Hearst N, Newman TB, Hulley SB. Delayed effects of the military draft on mortality. A randomized natural experiment. *N Engl J Med* 1986;314(10):620–624.
13. Tan HJ, Norton EC, Ye Z, et al. Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer. *JAMA* 2012;307(15):1629–1635.
14. Gum PA, Thamilarasan M, Watanabe J, et al. Aspirin use and all-cause mortality among patients being evaluated for known or suspected coronary artery disease: A propensity analysis. *JAMA* 2001;286(10):1187–1194.
15. Klungel OH, Martens EP, Psaty BM, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol* 2004;57(12):1223–1231.
16. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight “paradox” uncovered? *Am J Epidemiol* 2006;164(11):1115–1120.
17. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. *BMC Med* 2010;8:17.
18. Katz MH. *Multivariable analysis: a practical guide for clinicians*, 2nd ed. Cambridge, UK; New York: Cambridge University Press, 2006.

---

[1]Questões semelhantes surgem em estudos sobre testes diagnósticos (Capítulo 12), mas nessas situações o objetivo não é determinar um efeito causal, mas sim determinar se o teste que está sendo estudado acrescenta poder preditivo substancial para as informações já disponíveis no momento em que o estudo foi feito.

[2]Um bioestatístico que trabalha conosco brincou que elaborar um programa de análise estatística amigável e intuitivo é como construir um carro de forma que uma criança consiga alcançar o pedal.

[3] Essa descrição de análise de covariância (ANCOVA) é simplificada. Na verdade, a escolaridade dos pais é plotada contra o QI das crianças com níveis normais de chumbo e das crianças com níveis elevados de chumbo. Usa-se então a inclinação que oferece o melhor ajuste para o conjunto de dados. O modelo para esse tipo de ajuste pressupõe, dessa forma, relações lineares entre escolaridade e QI em ambos os grupos, e que as inclinações das retas nos dois grupos sejam iguais.

CAPÍTULO 10

Delineando um ensaio clínico randomizado cego

Steven R. Cummings, Deborah Grady e Stephen B. Hulley

Em um ensaio clínico, o investigador aplica uma **intervenção** e observa os seus efeitos sobre um ou mais **desfechos**. A principal vantagem de um ensaio clínico em relação a um estudo observacional é sua capacidade de **demonstrar causalidade**. A **alocação aleatória** da intervenção minimiza a influência de variáveis confundidoras, e o **cegamento** pode minimizar a possibilidade de os efeitos aparentes serem explicados por diferenças no uso de cointervenções nos grupos de tratamento e controle ou por vieses na mensuração ou adjudicação dos desfechos.

Entretanto, um ensaio clínico geralmente apresenta **custo elevado**, **demanda tempo**, aborda uma **questão restrita** e às vezes expõe os participantes a **potenciais danos**. Por esses motivos, devem ser reservados para **questões de pesquisa já amadurecidas** e quando estudos observacionais e outras linhas de evidência sugerem que uma intervenção possa ser eficaz, mas faltam evidências mais fortes para que ela possa ser aprovada ou recomendada. Nem todas as questões de pesquisa são passíveis de serem estudadas por um delineamento como o ensaio clínico. Por exemplo, não é factível estudar se o tratamento medicamentoso para colesterol-LDL elevado em crianças previne infarto muitas décadas depois e não é ético randomizar as pessoas para fumar cigarros reais ou falsos para determinar o efeito sobre o câncer de pulmão. Mesmo assim, sempre que possível, devem-se obter evidências de ensaios clínicos sobre a eficácia e segurança das intervenções clínicas.

Este capítulo enfoca o **delineamento** do já clássico **ensaio clínico randomizado cego**: como escolher a **intervenção** e o **controle**, definir os

desfechos e efeitos adversos, selecionar os **participantes**, aferir as **variáveis basais e de desfecho** e avaliar as abordagens para a **randomização** e o **cegamento**. No próximo capítulo, trataremos de delineamentos alternativos para ensaios clínicos e de temas relacionados à implementação e à análise estatística.

SELECIONANDO A INTERVENÇÃO E O CONTROLE

O ensaio clínico randomizado clássico é um **delineamento em paralelo e intergrupos** que inclui um grupo que recebe uma intervenção a ser testada e um grupo controle que recebe um tratamento não ativo (de preferência um placebo) ou um tratamento de comparação. O investigador aplica a intervenção e o controle, segue ambos os grupos ao longo do tempo e compara o desfecho entre os grupos de intervenção e controle (Figura 10.1).

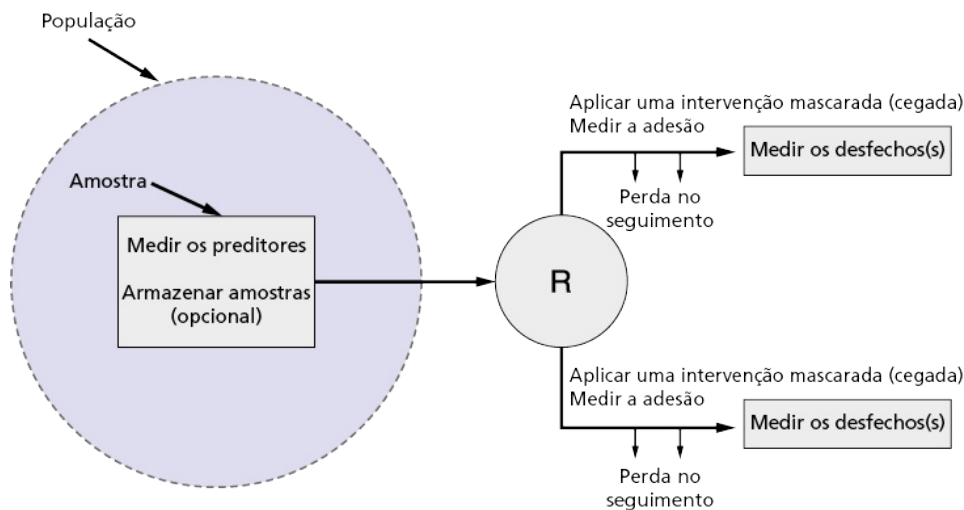



FIGURA 10.1 Em um ensaio clínico randomizado cego, os passo são:

- Selecionar uma amostra de sujeitos de uma população adequada para receber a intervenção.
- Medir as variáveis preditoras e, se adequado, os níveis basais da variável de desfecho. Considerar a opção de armazenar soro, imagens, e assim por diante, para análise posterior.
- Alocar aleatoriamente de forma mascarada a intervenção e o controle (p.ex., um placebo).
- Seguir a coorte ao longo do tempo, minimizando a perda no seguimento

e avaliando a adesão à intervenção e ao controle.

- Medir as variáveis de desfecho.

Escolha da intervenção

A escolha da intervenção é o primeiro passo crítico no delineamento de um ensaio clínico. Vários aspectos devem ser levados em conta nesse processo, incluindo a dosagem, a duração e a frequência da intervenção que proporcionam o melhor equilíbrio entre eficácia e segurança. Também é importante considerar a factibilidade do cegamento, o uso de uma ou mais intervenções, a aceitabilidade para os participantes e a capacidade de generalização para a forma como o tratamento será usado na prática clínica. Se houver incerteza sobre decisões importantes, como a dose que proporciona o melhor equilíbrio entre eficácia e segurança, é geralmente mais adequado postergar ensaios clínicos de grande porte e custo elevado até que estudos preliminares tenham sido concluídos para ajudar a esclarecer essas incertezas.

O equilíbrio entre **eficácia** e **segurança** depende da intervenção e da doença que estão sendo estudadas. A eficácia é mais importante quando se planejam intervenções para o tratamento de doenças que causam sintomas graves ou morte. Nesse caso, a melhor alternativa pode ser escolher a "dose máxima tolerável", como é o caso do câncer metastático. A segurança deve ser o principal critério quando se planejam intervenções para tratar condições sintomáticas que raramente resultam em progressão da doença ou morte. O **tratamento preventivo** em pessoas saudáveis deve passar pelos testes mais rigorosos de segurança: se o tratamento for eficaz, ele impedirá que algumas pessoas desenvolvam a doença, porém todos os indivíduos tratados irão correr riscos de efeitos adversos do tratamento. Nesse caso, a melhor opção geralmente é escolher uma dose que maximize a eficácia com um risco muito baixo de efeitos adversos. Se não houver certeza sobre a melhor dose com base em estudos prévios em animais ou humanos, pode ser necessário realizar ensaios clínicos que comparem os efeitos de diferentes doses sobre marcadores intermediários ou desfechos clínicos (veja ensaios clínicos de fase II, Capítulo 11).

Às vezes, o investigador pode decidir comparar **várias doses** ou níveis de intensidade com um único grupo-controle. Por exemplo, quando se

planejou o estudo Multiple Outcomes of Raloxifene Evaluation Trial, não havia certeza sobre a melhor dose do raloxifeno (60 ou 120 mg), então, o estudo testou duas doses diferentes para a prevenção de fraturas vertebrais (1). Isso pode ser uma alternativa razoável, mas tem custos: um ensaio clínico maior e com custo mais elevado, além da complexidade de se lidar com hipóteses múltiplas (Capítulo 5).

Certos tratamentos costumam ter a dose ajustada de modo a otimizar o efeito para cada paciente. Nessas circunstâncias, o melhor a fazer é planejar uma intervenção que garanta que a **dose do medicamento ativo possa ser tateada** para alcançar um desfecho clínico-alvo, como a redução da carga viral de hepatite C. Para manter o cegamento, mudanças correspondentes também devem ser feitas na “dose” do placebo (por alguém não envolvido no ensaio clínico) para um participante do grupo-controle selecionado aleatoriamente ou pareado.

Ensaios clínicos que testam **intervenções isoladas** são geralmente mais fáceis de planejar e de implementar do que os que testam combinações de tratamentos. No entanto, **combinações** de tratamentos são usadas para tratar muitas condições clínicas, como infecção pelo HIV ou insuficiência cardíaca. A desvantagem principal de se testar combinações de tratamentos é que o resultado não poderá fornecer conclusões claras sobre nenhum dos elementos das intervenções. Em um dos ensaios clínicos do Women’s Health Initiative, por exemplo, mulheres pós-menopáusicas foram tratadas com estrogênio mais progestogênio ou placebo. A intervenção aumentou o risco de vários desfechos, incluindo câncer de mama, no entanto, ficou a dúvida sobre se o efeito se deveu ao estrogênio ou ao progestogênio (2). Em geral, é preferível delinear ensaios clínicos com apenas uma diferença importante entre dois grupos de estudo.

É importante avaliar o quão receptivos os participantes serão à intervenção proposta e se ela poderá ser mascarada. Deve-se considerar também se intervenção poderá ser incorporada na prática médica. **Intervenções simples** são geralmente melhores do que as mais complicadas (é mais fácil garantir que os pacientes tomem um comprimido uma vez ao dia do que injeções subcutâneas duas ou três vezes ao dia). Intervenções complicadas com aspectos qualitativos, como aconselhamento multifacetado de pacientes sobre mudanças no comportamento, podem não ser factíveis, por serem difíceis de replicar,

consumirem muito tempo ou apresentarem custo elevado. Tais intervenções tendem a ter um impacto menor em termos de saúde pública, mesmo que um ensaio clínico comprove que sejam eficazes.

Escolha do controle

O melhor grupo-controle é aquele que **não recebe tratamento ativo** e que, ao mesmo tempo, pode ser **cegado.** Isso, para medicamentos, significa receber um **placebo** idêntico ao tratamento ativo. Essa estratégia compensa para qualquer efeito-placebo da intervenção ativa (pela sugestão ou expectativa de benefício), de forma que qualquer diferença no desfecho possa ser atribuída a um efeito específico das intervenções.

A melhor comparação entre os grupos de intervenção e controle ocorre quando não há **cointervenções** – medicamentos, terapias ou comportamentos (que não a intervenção sob estudo) que alteram o risco de desenvolver o desfecho de interesse. Por exemplo, em um ensaio clínico randomizado sobre uma intervenção de yoga comparada ao cuidado convencional para prevenir diabetes, a equipe poderia estimular os participantes a fazerem mais exercício físico e a perderem peso. Essas são cointervenções potencialmente eficazes que reduzem o risco de desenvolver diabetes. Quando os participantes de ambos os grupos são submetidos a cointervenções eficazes, a taxa de desenvolvimento do desfecho é diminuída, o poder é reduzido e o tamanho de amostra ou a duração do estudo precisam ser maiores. Quando o uso de cointervenções eficazes difere entre os grupos de intervenção e controle, o desfecho é enviesado. Na ausência de um bom cegamento, o protocolo deve incluir planos para obtenção de dados que permitam ajuste estatístico para diferenças entre os grupos na taxa de uso dessas cointervenções. No entanto, medir as cointervenções pode ser difícil e o ajuste para essas diferenças após a randomização deve ser visto como uma análise secundária ou explanatória, uma vez que viola o princípio da intenção de tratar (Capítulo 11).

Nem sempre é possível deixar de dar qualquer tipo de tratamento que não seja a intervenção do estudo. Por exemplo, em um ensaio clínico para saber se um novo medicamento reduz o risco de infarto do miocárdio em pacientes com doença coronariana conhecida, não seria ético proibir ou desencorajar os participantes a usarem tratamentos médicos indicados

para essa doença, como aspirina, estatinas e β-bloqueadores. Uma opção seria prescrever um **tratamento-padrão para todos os participantes** do estudo. Embora essa estratégia reduza a taxa global de eventos, aumentando assim o tamanho de amostra, ela testa a questão clínica mais relevante: se a nova intervenção melhora o desfecho quando aplicada junto com o tratamento-padrão.

Quando o tratamento a ser estudado é um novo medicamento que se acredita ser uma boa alternativa ao tratamento-padrão, uma opção é delinear um **ensaio clínico de não inferioridade** ou um **ensaio clínico de equivalência**, que compara o novo tratamento com um tratamento já comprovadamente eficaz (veja Capítulo 11).

ESCOLHENDO MEDIDAS DE DESFECHO

A definição dos desfechos a serem estudados influencia muitos outros componentes do delineamento. Influencia também o custo e a factibilidade do ensaio clínico. Devem-se, em geral, incluir vários desfechos, com o objetivo de aumentar a riqueza dos resultados e as possibilidades de análises secundárias. No entanto, um desses desfechos deve ser designado como **desfecho primário** (ou **desfecho principal**), refletindo a questão de pesquisa principal, permitindo o cálculo do tamanho de amostra e estabelecendo as prioridades na hora de implementar o estudo.

Desfechos clínicos fornecem as melhores evidências para embasar a adoção de um tratamento ou intervenção preventiva e para orientar como ele deverá ser empregado. No entanto, para desfechos raros, como o desenvolvimento de câncer, geralmente são necessários ensaios clínicos de grande porte, longa duração e custo elevado. Como discutido no Capítulo 6, desfechos medidos como variáveis contínuas – por exemplo, qualidade de vida – podem geralmente ser estudados com um número menor de participantes do que quando o desfecho é dicotômico. Entretanto, o desfecho clínico mais importante muitas vezes será obrigatoriamente dicotômico, como recorrência de câncer de mama.

Marcadores intermediários, como densidade óssea, são medidas relacionadas ao desfecho clínico. Ensaios clínicos que usam desfechos intermediários podem aprofundar nosso conhecimento sobre a

fisiopatologia e fornecer informações para escolher a melhor dose ou frequência de tratamento em estudos com desfechos clínicos. A relevância clínica de estudos com desfechos intermediários depende de até que ponto mudanças nesses marcadores, especialmente aquelas decorrentes do tratamento, representam mudanças no risco de desfechos clínicos. Marcadores intermediários podem ser considerados **marcadores substitutos** para o desfecho clínico se as mudanças induzidas pelo tratamento no marcador puderem predizer de forma consistente como o tratamento irá influenciar o desfecho clínico (3). Em geral, um bom marcador substituto mede alterações em um fator intermediário na rota causal que determina o desfecho clínico.

A carga viral do HIV é um bom marcador substituto, porque tratamentos que reduzem a carga viral consistentemente reduzem a morbidade e a mortalidade em pacientes com infecção pelo HIV. Por outro lado, a densidade mineral óssea é um marcador substituto ruim (3). Ela reflete a quantidade de conteúdo mineral em um corte de osso, mas tratamentos que aumentam a densidade mineral óssea têm pouco ou nenhum efeito sobre o risco de fraturas. A magnitude do aumento nessa variável intermediária não está consistentemente relacionada com o quanto o tratamento reduz o risco de fratura (4). A melhor evidência de que um marcador biológico é um bom desfecho substituto vem de ensaios clínicos randomizados com desfechos clínicos (p. ex., *fraturas*) que também medem em todos os participantes mudanças no potencial marcador intermediário (p. ex., *densidade mineral óssea*). Se o marcador for um bom substituto, então ajustes estatísticos para mudanças no marcador representarão grande parte do efeito do tratamento sobre o desfecho (3).

Número de variáveis de desfecho

Muitas vezes é desejável ter um **grande número de variáveis de desfecho** que meçam diferentes aspectos dos fenômenos de interesse. No Heart and Estrogen/Progestin Replacement Study (HERS), eventos coronarianos foram escolhidos como desfecho principal. A revascularização miocárdica, a hospitalização por angina instável ou insuficiência cardíaca, o acidente vascular encefálico e o ataque isquêmico transitório, a tromboembolia venosa e a mortalidade por todas

as causas foram também avaliados e adjudicados para fornecer uma descrição mais detalhada dos efeitos cardiovasculares da terapia hormonal (5). No entanto, um **único desfecho principal** (*eventos coronarianos*) foi designado para fins de planejamento do tamanho de amostra e da duração do estudo e para evitar os problemas relacionados à interpretação de testes de hipóteses múltiplas (Capítulo 5).

Desfechos compostos

Alguns ensaios clínicos definem desfechos que são compostos por uma série de diferentes eventos ou medidas. Por exemplo, muitos ensaios clínicos sobre intervenções para reduzir o risco de doença coronariana incluem em uma única variável de desfecho diversos eventos coronarianos específicos, como infarto do miocárdio, morte coronariana e procedimentos de revascularização miocárdica. Isso pode ser razoável se cada um desses desfechos for clinicamente importante, se o tratamento de cada uma das condições funcionar por mecanismos semelhantes e se a intervenção puder reduzir o risco de cada um dos eventos. Além disso, um desfecho composto geralmente fornece poder estatístico maior do que um desfecho único, pois ele resulta em um número maior de eventos. Entretanto, desfechos compostos que incluem eventos que não são tão relevantes clinicamente ou que ocorrem com uma frequência muito maior do que os demais podem resultar em achados enganadores. Por exemplo, se a hospitalização para avaliação de dor torácica for incluída no desfecho coronariano composto, esse evento irá dominar os demais se a hospitalização ocorrer com uma frequência muito maior do que o infarto do miocárdio, a morte coronariana ou a revascularização. Assim, quando uma intervenção altera o desfecho composto, poder-se-ia dizer que ela reduz o risco de "eventos coronarianos" quando, na verdade, ela apenas reduz o risco de hospitalização por dor torácica.

Os desfechos compostos devem ser planejados cuidadosamente. Se o tratamento produzir apenas um efeito pequeno sobre um desfecho, especialmente se esse desfecho for relativamente comum, ele pode acrescentar pouco poder estatístico ou até mesmo aumentar o tamanho de amostra necessário para detectar um efeito. Por exemplo, se o acidente vascular encefálico for acrescentado a um desfecho cardiovascular composto, a intervenção poderia reduzir o risco de eventos coronarianos,

não ter impacto ou mesmo aumentar o risco de acidente vascular encefálico e, portanto, mostrar ausência de efeito sobre o desfecho cardiovascular composto.

Efeitos adversos

O investigador deve incluir medidas de desfecho para detectar possíveis **efeitos adversos** resultantes da intervenção. A maioria dos ensaios clínicos tem, entre seus objetivos principais, avaliar se os benefícios de uma intervenção superam seus efeitos adversos, mesmo quando testam tratamentos aparentemente inócuos como um programa de educação em saúde. Os efeitos adversos podem variar desde sintomas relativamente leves, como uma erupção cutânea leve e transitória, a complicações sérias e fatais. A taxa de ocorrência, o efeito do tratamento e as exigências de tamanho de amostra para detectar efeitos adversos são geralmente diferentes daqueles necessários para detectar benefícios. Infelizmente, efeitos adversos raros costumam ser impossíveis de detectar, mesmo em ensaios clínicos de grande porte, e são apenas descobertos (se forem descobertos) por grandes estudos observacionais ou por relatos de caso, após uma intervenção já estar disseminada na prática clínica.

Nos estágios iniciais dos testes de um novo tratamento, quando ainda não há clareza sobre seus potenciais efeitos adversos, os investigadores devem fazer perguntas amplas e abertas sobre todos os tipos de efeitos adversos que poderão ocorrer. Em ensaios clínicos de grande porte, a avaliação e a codificação de todos os potenciais eventos adversos pode ser cara e consumir muito tempo, muitas vezes trazendo poucos resultados importantes. É importante considerar estratégias para minimizar esse problema e ao mesmo tempo preservar uma avaliação adequada dos potenciais malefícios da intervenção. Por exemplo, em ensaios clínicos muito grandes, eventos comuns e pouco importantes, como infecções de vias aéreas superiores e desconforto gastrintestinal, podem ser registrados em um subconjunto dos participantes. Pode não ser necessário registrar efeitos adversos que não sejam graves se estudos prévios não mostraram evidências na incidência de sintomas menores. Além dessas questões abertas, perguntas específicas devem ser elaboradas para identificar eventos adversos esperados em função de estudos prévios ou da experiência clínica. Por exemplo, visto que a miosite é um efeito adverso

já conhecido das estatinas, os sinais e sintomas da miosite devem ser investigados em qualquer ensaio clínico sobre uma nova estatina.

Os efeitos adversos relatados como sintomas ou termos clínicos devem ser codificados e categorizados para que possam ser analisados. Alguns dicionários de termos que são frequentemente utilizados são o MedDRA (www.ich.org/products/meddra.html) e o SNOMED (https://www.nlm.nih.gov/research/umls/). Neles, os termos são agrupados de diversas formas, como, por exemplo, por sintomas, diagnósticos específicos e órgãos ou sistemas afetados. Por exemplo, um evento adverso registrado como "febre e tosse" e um evento adverso registrado como "bronquite" serão agrupados junto com outras condições, como pneumonia, sob o verbete "infecção respiratória" e, em uma hierarquia superior, como um efeito adverso no sistema respiratório. Esses esquemas de classificação fornecem um bom resumo dos efeitos adversos e são razoavelmente acurados para doenças que permitem um diagnóstico específico, como fraturas. Entretanto, podem deixar passar eventos adversos que são descritos por vários termos diferentes se esses termos não estiverem agrupados. Por exemplo, em um ensaio clínico sobre o denosumabe para a prevenção de fraturas osteoporóticas, o MedDRA codificou casos de celulite de forma separada dos casos de erisipela (dois nomes para o mesmo tipo de infecção). Quando combinados, 12 casos graves de celulite ocorreram com o denosumabe versos 1 com placebo ($P < 0{,}001$) (6).

Os efeitos adversos são também geralmente classificados de acordo com a gravidade. Eventos adversos graves (EAGs) são definidos como morte ou eventos que ameaçam a vida, eventos que requerem ou prolongam a internação, incapacidade ou dano permanente, malformações congênitas e outros eventos médicos importantes que podem requerer intervenção médica ou cirúrgica para prevenir um dos outros desfechos (www.fda.gov/Safety/MedWatch/HowToReport/ucm053087.htm). Eventos adversos graves geralmente devem ser prontamente relatados para os comitês de ética em pesquisa e para a instituição que financia o estudo.

Quando os dados de um ensaio clínico são usados para solicitar a aprovação de um novo medicamento, o delineamento do estudo deve atender às exigências das autoridades reguladoras para o relato de eventos

adversos
(http://www.fda.gov/Drugs/InformationOnDrugs/ucm135151.htm).
Algumas áreas da medicina, como a oncologia, apresentam métodos já estabelecidos para a classificação dos eventos adversos
(http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.h

SELECIONANDO OS PARTICIPANTES

O Capítulo 3 discutiu como especificar critérios de entrada, definindo uma população-alvo que seja apropriada para a questão de pesquisa e uma população acessível que seja prática de se estudar; discutiu também como delinear uma estratégia eficiente e científica para selecionar os participantes e como recrutar os sujeitos especificados para estudo. A seguir, abordamos alguns pontos adicionais de relevância especial para ensaios clínicos.

Definir os critérios de entrada

Em um ensaio clínico, os critérios de inclusão e exclusão têm como objetivo identificar uma população para a qual seja factível, ético e relevante estudar o impacto da intervenção sobre os desfechos. Os **critérios de inclusão** devem permitir o arrolamento de um número suficiente de participantes com uma taxa suficientemente alta do desfecho principal, possibilitando um poder adequado para encontrar um efeito importante da intervenção sobre o desfecho. Por outro lado, os critérios devem também maximizar a capacidade de generalização dos achados do estudo e a facilidade de recrutamento. Por exemplo, quando o desfecho de interesse é um evento incomum, como câncer de mama, é geralmente necessário recrutar participantes com alto risco para o desfecho, com o objetivo de reduzir o tamanho de amostra e o tempo de seguimento em níveis factíveis. Por outro lado, estreitar os critérios de inclusão para indivíduos de alto risco pode limitar a capacidade de generalização dos resultados e dificultar o recrutamento de um número suficiente de participantes.

Para planejar um **tamanho de amostra** adequado, devem-se obter estimativas confiáveis da taxa do desfecho principal nos indivíduos potencialmente elegíveis. Essas estimativas podem basear-se em dados de

estatísticas vitais, em dados de estudos observacionais longitudinais ou em taxas observadas no grupo não tratado em ensaios clínicos com critérios de inclusão semelhantes. Por exemplo, a taxa esperada de câncer de pâncreas em adultos pode ser estimada a partir de registros de câncer. No entanto, o investigador deve estar atento para os efeitos do rastreamento e do uso de voluntários saudáveis, que fazem com que as taxas de eventos entre aqueles que atendam os critérios de seleção e concordem em participar sejam menores que as encontradas na população geral. Pode ser preferível obter taxas de câncer de pâncreas do grupo-placebo de outros ensaios clínicos com critérios de seleção semelhantes.

Incluir pessoas com **alto risco** para o desfecho pode diminuir o número de participantes necessários para o estudo. Se o desfecho de interesse já tiver fatores de risco estabelecidos, é possível planejar os critérios de seleção de forma a incluir participantes com um risco estimado mínimo para o desfecho. O ensaio clínico Raloxifene Use for The Heart, que estudou o efeito do raloxifeno na prevenção de doença cardiovascular e do câncer de mama, selecionou mulheres com maior risco de doença cardiovascular com base em uma combinação de fatores de risco (7). Outra forma de aumentar a taxa de eventos é limitar o arrolamento a indivíduos que já têm a doença. O Heart and Estrogen/Progestin Replacement Study (HERS) incluiu 2.763 mulheres que já tinham doença coronariana para testar se estrogênio mais progestogênio reduziria o risco de novos eventos coronarianos (5). Essa estratégia apresentou um custo muito menor do que o ensaio clínico do Women’s Health Initiative sobre a mesma questão de pesquisa em mulheres sem doença coronariana, que precisou arrolar aproximadamente 17 mil participantes (8).

Apesar das vantagens das **amostras probabilísticas de populações gerais** em estudos observacionais, para a maioria dos ensaios clínicos randomizados esse tipo de amostragem não é factível nem necessária. É verdade que incluir participantes com características diversas aumenta a validade externa dos resultados para um contexto mais amplo. No entanto, a não ser que haja diferenças biológicas ou genéticas entre as populações que influenciam o efeito terapêutico, os resultados de um ensaio clínico conduzido em uma amostra de conveniência (p. ex., mulheres com doença coronariana que responderam a anúncios) são geralmente semelhantes aos obtidos em amostras probabilísticas de indivíduos elegíveis (todas as

mulheres com doença coronariana). Em algumas ocasiões, a eficácia do tratamento depende de características dos sujeitos, o que recebe o nome de modificação de efeito ou interação (ver Capítulo 11). Por exemplo, alguns tratamentos para osteoporose reduzem de forma substancial o risco de fratura em mulheres com densidade óssea muito baixa (escores-T inferiores a -2,5), com pouco ou nenhum efeito em mulheres com densidade óssea mais elevada (P = 0,02 para interação) (9, 10). Nesse caso, incluir em um ensaio clínico apenas mulheres com densidade óssea muito baixa pode aumentar a magnitude do efeito e reduzir o tamanho de amostra em um ensaio clínico sobre esse tipo de tratamento.

Estratificar os participantes de acordo com uma característica, como grupo racial, permite aos investigadores arrolar um número desejado de participantes com uma característica que pode influenciar o efeito terapêutico ou a capacidade de generalização. O recrutamento para um determinado estrato pode ser encerrado quando a meta de número de participantes com aquela característica foi alcançada. Entretanto, como a maioria dos ensaios clínicos não tem poder estatístico suficiente para testar para a heterogeneidade nos efeitos da intervenção entre os subgrupos, essa estratégia pode ter valor prático limitado.

Os **critérios de exclusão** devem ser parcimoniosos, pois exclusões desnecessárias podem dificultar o recrutamento do número necessário de participantes, reduzir a capacidade de generalização dos resultados e aumentar a complexidade e o custo do recrutamento. Há cinco motivos principais para excluir indivíduos de um ensaio clínico (Tabela 10.1).

TABELA 10.1 Motivos para excluir indivíduos de um ensaio clínico

MOTIVO	EXEMPLO (ENSAIO CLÍNICO SOBRE RALOXIFENO VS. PLACEBO NA PREVENÇÃO DE DOENÇA CORONARIANA)
1. Um dos tratamentos do estudo pode ser prejudicial	
• Risco inaceitável de reação adversa ao tratamento ativo	Tromboembolia venosa prévia (o raloxifeno aumenta o risco desse evento)
• Risco inaceitável de alocação a placebo	Câncer de mama positivo para receptor de estrogênio recente (o tratamento com antiestrogênio é um padrão de tratamento eficaz)
2. O tratamento ativo não pode ser eficaz ou é improvável que o seja, pois o indivíduo	
• Apresenta baixo risco para o desfecho	Mulheres adolescentes, com risco muito baixo para doença

	coronariana
• Tem um tipo de doença que provavelmente não irá responder ao tratamento	Pacientes com doença cardíaca valvular, que provavelmente não irão responder aos efeitos antiaterogênicos hipotetizados do raloxifeno
• Já faz tratamento que provavelmente irá interferir na intervenção	Já toma estrogênios (que competem com o raloxifeno)
3. Baixa probabilidade de aderir à intervenção	Baixa adesão durante o período de *run-in* (Capítulo 11)
4. Baixa probabilidade de completar o período de seguimento	Planeja se mudar antes do término do estudo e não estará disponível para as aferições finais do desfecho Baixa expectativa de vida devido à doença grave
5. Problemas de ordem prática para participação no protocolo	Deficiência mental que impossibilita respostas corretas

Os potenciais participantes deveriam ser excluídos se o tratamento ou a intervenção controle **não forem seguros** para eles. O tratamento ativo pode não ser seguro em indivíduos suscetíveis a efeitos adversos, conhecidos ou suspeitos, do tratamento ativo. Por exemplo, o infarto do miocárdio é um efeito adverso raro do tratamento com sildenafil (Viagra®), de modo que ensaios clínicos sobre esse fármaco no tratamento do vasoespasmo doloroso em pacientes com doença de Raynaud devem excluir pacientes com doença coronariana (11). Por outro lado, ser designado para o grupo de tratamento inativo ou placebo pode não ser seguro para certos participantes. Por exemplo, em mulheres com fraturas vertebrais, os bisfosfonatos comprovadamente reduzem o risco de fraturas subsequentes, tornando inaceitável incluí-las em um ensaio clínico controlado por placebo sobre um novo tratamento para osteoporose, a não ser que os bisfosfonatos também fossem fornecidos a todos os participantes. Os sujeitos para os quais é pouco provável que o tratamento ativo seja eficaz também devem ser excluídos, assim como aqueles que provavelmente não irão aderir à intervenção ou que poderão não completar o período de seguimento. Eventualmente, problemas de ordem prática, como deficiência mental, que torne difícil seguir instruções, também justificam exclusão do estudo. Deve-se pesar cuidadosamente os potenciais critérios de exclusão que se apliquem a muitos indivíduos (p. ex., *diabetes ou limites superiores de idade*), pois eles terão grande impacto na factibilidade e nos custos do recrutamento, assim como na capacidade de generalização dos resultados.

Delinear um tamanho de amostra adequado e planejar com base nele o recrutamento

Ensaios clínicos com número insuficiente de participantes para detectar efeitos importantes geram gastos desnecessários, são antiéticos e podem levar a conclusões enganadoras (12). Estimar o tamanho de amostra é, pois, uma das partes mais importantes do início da fase de planejamento de um ensaio clínico (Capítulo 6). Ao fazê-lo, deve-se levar em consideração o fato de que as taxas do desfecho em ensaios clínicos são geralmente mais baixas do que o estimado, devido ao viés do participante voluntário saudável. Além disso, muitas vezes é mais difícil recrutar participantes para um ensaio clínico do que para um estudo observacional, pois eles devem estar dispostos a serem randomizados, muitas vezes a um placebo ou a um medicamento "experimental". Por esses motivos, é importante planejar uma amostra com tamanho expressivo de uma população grande e acessível, bem como prever tempo e recursos suficientes para recrutar o tamanho de amostra desejado quando as barreiras para alcançá-lo se tornam maiores que o previsto (o que frequentemente é o caso).

MEDINDO VARIÁVEIS BASAIS

Para facilitar o contato com participantes perdidos no seguimento, é importante registrar os nomes, números de telefone, endereços e e-mails de dois ou três amigos ou parentes que poderão ajudar a localizar o participante. Se for permitido, também vale a pena registrar o número do CPF ou de outros documentos importantes que poderão ser usados para averiguar o estado vital dos participantes (p. ex, nos Estados Unidos, por meio do National Death Index) ou a ocorrência de desfechos importantes por meio de registros de saúde (p. ex., bases de dados de sistemas de seguro de saúde). No entanto, essas são "informações de saúde pessoal protegidas", ou seja, informações confidenciais que não devem acompanhar os dados enviados ao centro de coordenação ou à instituição financiadora.

Descrever os participantes

É importante coletar informações sobre fatores de risco estabelecidos ou

potenciais para o desfecho e sobre características dos participantes que poderão afetar a eficácia da intervenção. Essas medidas permitem também verificar a comparabilidade na linha de base dos grupos randomizados e fornecem informações para avaliar a capacidade de generalização dos achados. A meta é garantir que as diferenças nas características basais não excedam o que poderia ser esperado pelo acaso; se as diferenças excederem o acaso, pode ter havido erro técnico ou viés no processo de randomização. Em ensaios clínicos de pequeno porte, que são suscetíveis a distribuições desiguais substanciais das características basais entre os grupos randomizados, resultantes do acaso, medir preditores importantes do desfecho permite ajustar estatisticamente para a comparação randomizada, o que reduz os efeitos da distribuição desigual. Essa medição também permite examinar se a intervenção afeta de forma desigual **subgrupos** classificados de acordo com variáveis basais (**modificação de efeito**, ver o Capítulo 11).

Medir a variável de desfecho na linha de base

Se os desfechos incluírem mudança em uma determinada variável, esta deve ser aferida no início do estudo da mesma forma que será aferida no final. Quando a variável de desfecho é contínua (p. ex., efeitos da terapia cognitivo-comportamental sobre escores de depressão), a melhor medida é geralmente uma mudança ao longo da duração do estudo. Essa abordagem em geral minimiza a variabilidade no desfecho entre os participantes do estudo e oferece maior poder estatístico do que simplesmente comparar valores no final do ensaio clínico. Em estudos com uma variável de desfecho dicotômica (p. ex., incidência de doença coronariana), pode ser importante demonstrar, por meio da história clínica e do eletrocardiograma, que a doença não estava presente na linha de base. Pode também ser útil medir na linha de base variáveis de desfecho secundárias e desfechos de estudos suplementares (*ancillary studies*) já planejados.

Ser parcimonioso

Tendo apontado os múltiplos usos para as medidas basais, deve-se enfatizar que o delineamento básico de um ensaio clínico não exige que essas medições sejam feitas, pois a randomização minimiza o problema

do confundimento por fatores presentes no início do estudo. Fazer muitas medições aumenta os custos e a complexidade do estudo. Em ensaios clínicos randomizados com orçamento limitado, o tempo e os recursos podem ser mais bem aplicados em aspectos vitais à integridade do estudo, como a suficiência do tamanho de amostra, o sucesso da randomização e do cegamento, e a completude da adesão e do seguimento. Com base nisso, Yusuf e colaboradores vêm promovendo o uso de grandes ensaios clínicos com um número muito pequeno de aferições (13).

Bancos de imagens e de amostras biológicas

Armazenar imagens, soro, DNA, etc. na linha de base permite medir posteriormente as mudanças decorrentes do tratamento, marcadores que predizem o desfecho e fatores, como genótipo, que poderão identificar indivíduos que respondem bem ou mal ao tratamento. Materiais biológicos armazenados também podem ser uma fonte rica para estudar outras questões de pesquisa não diretamente relacionadas ao desfecho principal.

RANDOMIZANDO E CEGANDO

O quarto passo na Figura 10.1 é alocar aleatoriamente os participantes em dois grupos. No delineamento mais simples, um grupo recebe um tratamento ativo e o outro recebe placebo. A **alocação aleatória** garante que fatores como idade, sexo e outras características prognósticas basais que confundiriam uma associação observada (incluindo fatores desconhecidos ou não aferidos) sejam distribuídos igualmente entre os grupos randomizados na linha de base, exceto pela variação ao acaso. O **cegamento** é importante para manter a comparabilidade dos grupos durante o ensaio clínico e para assegurar uma avaliação dos desfechos livre de viés.

Randomização

Como a randomização é a base de um ensaio clínico, é importante que seja feita de forma correta. As duas características mais importantes da randomização são que ela deve realmente **alocar tratamentos aleatoriamente** e que as alocações devem ser **invioláveis**, de forma que

seja impossível que fatores intencionais ou não intencionais influenciem na randomização.

Antes da randomização, é importante que o participante complete a coleta de dados da linha de base, seja considerado elegível para inclusão e consinta em ingressar no estudo. Então, ele será designado aleatoriamente a um grupo, usando um algoritmo computadorizado ou aplicando-se uma série de números aleatórios. A partir do momento em que se gera uma lista da ordem aleatória de alocação, essa lista deve ser aplicada aos participantes à medida que eles vão ingressando no estudo.

É fundamental planejar o procedimento de alocação aleatória de forma que os membros da equipe de pesquisa não possam influenciar a alocação. Por exemplo, em ensaios clínicos realizados em um único centro, as alocações aleatórias para tratamento podem ser colocadas *a priori* em envelopes lacrados por alguém que não estará envolvido na abertura dos envelopes. Os envelopes devem ser numerados de forma que todos possam ser contabilizados ao final do estudo; opacos, para prevenir a transluminação por luz forte; e invioláveis. Quando um participante é randomizado, o seu nome e o número do próximo envelope não aberto são registrados na *presença de um segundo membro da equipe,* e ambos assinam seus nomes no envelope; *então,* o envelope é aberto e o grupo de tratamento contido nele é alocado ao participante e registrado por escrito.

Ensaios clínicos multicêntricos geralmente têm um setor de randomização inviolável separado que os membros do estudo podem contatar quando um participante elegível está pronto para ser randomizado. O membro da equipe fornece o nome e o número do novo participante no estudo. Essa informação é registrada, e o grupo de tratamento é então aleatoriamente designado com base em um programa de computador que fornece um número de alocação do tratamento associado às intervenções do estudo. O tratamento também pode ser aleatoriamente designado por programas de computador no local da pesquisa, desde que esses programas sejam invioláveis. Deve-se tomar muito cuidado para prevenir a violação da randomização, pois os investigadores, às vezes, podem se sentir pressionados a interferir na alocação do tratamento (p. ex., quando um indivíduo parece se adequar mais ao grupo de tratamento ativo em um ensaio clínico controlado por placebo).

Considerar técnicas especiais de randomização

Em geral, a abordagem preferencial é fazer uma randomização simples dos participantes individuais para cada grupo de intervenção. Em ensaios clínicos de pequeno ou médio porte, é possível aumentar levemente o poder estatístico por meio de procedimentos especiais de randomização para balancear o número de participantes em cada grupo (randomização em blocos) e a distribuição das variáveis basais sabidamente preditoras do desfecho (randomização em blocos e estratificada).

A **randomização em blocos** é uma técnica comumente usada para garantir que o número de participantes seja igualmente distribuído entre os grupos do estudo. A randomização é feita em "blocos" de tamanhos predeterminados. Por exemplo, se o bloco contiver seis pessoas, a randomização é feita normalmente dentro de cada bloco de seis, até que três pessoas tenham sido randomizadas para um dos grupos; então, os indivíduos seguintes são automaticamente designados ao outro grupo, até que o bloco de seis seja completado. Isso significa que, em um estudo com 30 participantes, exatamente 15 serão alocados para cada grupo e, em um estudo com 33 participantes, a desproporção não pode ser maior que 18:15. A randomização em blocos de tamanho fixo pode não funcionar muito bem em estudos não cegos, pois a designação do tratamento para os participantes no final de cada bloco poderia ser antecipada e manipulada. Esse problema pode ser minimizado variando-se o tamanho dos blocos aleatoriamente (variando, p. ex., de blocos de 4 a 8) de acordo com um esquema desconhecido pelo investigador.

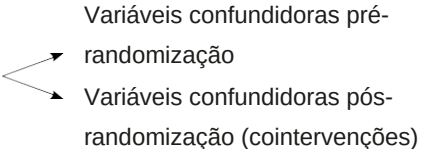
A **randomização em blocos e estratificada** garante que um preditor importante do desfecho seja distribuído mais uniformemente entre os grupos de estudo do que o mero acaso poderia determinar. Em um ensaio clínico sobre o efeito de um medicamento na prevenção de fraturas, a fratura vertebral prévia é um preditor tão forte do desfecho que talvez a melhor coisa a fazer seja garantir que números semelhantes de indivíduos com fraturas vertebrais sejam alocados em cada grupo. Isso pode ser alcançado realizando-se a randomização em blocos separadamente por "estratos" – os com e os sem fraturas vertebrais. A randomização em blocos e estratificada pode aumentar ligeiramente o poder de um ensaio clínico de pequeno porte, pois reduz a variação do desfecho causada por

desproporções ao acaso de preditores basais importantes. Esse tipo de randomização traz poucos benefícios em ensaios clínicos de grande porte (mais de mil participantes), pois a designação aleatória garante uma distribuição quase parelha das variáveis basais.

Uma limitação importante da randomização em blocos e estratificada é o número pequeno de variáveis basais (não mais do que duas ou três) que podem ser balanceadas por essa técnica. Uma técnica para contornar essa limitação é a **randomização adaptativa**, que usa uma “moeda enviesada” para alterar a probabilidade de designar cada novo participante, de modo que, por exemplo, um indivíduo com um escore de risco elevado com base em qualquer número de variáveis prognósticas basais teria uma probabilidade ligeiramente maior de integrar o grupo de estudo que tem menor risco geral com base em todos os participantes já randomizados até aquele momento. As desvantagens dessa técnica incluem a dificuldade de explicar aos participantes, durante o consentimento informado, a probabilidade de alocação aos grupos de estudo, bem como a complexidade de implementação, por requerer um sistema computadorizado interativo que recomputa as probabilidades da moeda enviesada a cada nova randomização.

Geralmente, é melhor randomizar números iguais para cada grupo, pois isso maximiza o poder estatístico para um determinado tamanho total de amostra. Entretanto, a atenuação do poder mesmo com uma desproporção de 2:1 é relativamente modesta (14), sendo às vezes adequado **alocar de forma desigual** os participantes para os grupos de tratamento e controle (15). Isso pode ocorrer das seguintes formas:

- Aumentando a razão de tratamento ativo contra tratamento controle, de modo a tornar o estudo mais **atraente** para potenciais participantes, como aqueles infectados pelo HIV, que iriam apreciar a maior probabilidade de receber tratamento ativo se arrolados.
- Diminuindo a razão de participantes com tratamento ativo contra participantes com tratamento controle, de modo a tornar o estudo mais **acessível** quando as intervenções tiverem custo muito elevado (como no estudo sobre dieta com baixo teor de gorduras do Women’s Health Initiative (16)).

- Aumentando a proporção designada para o grupo que serve como controle para diversos grupos de tratamento ativo **aumenta o poder** de cada comparação por aumentar a precisão da estimativa do grupo controle (como ocorreu no estudo Coronary Drug Project (17)).

A **randomização pareada** é uma estratégia para balancear confundidores na linha de base que exige que sejam selecionados pares de participantes de acordo com características importantes como idade e sexo, então alocando aleatoriamente um membro de cada par para cada grupo de estudo. Uma desvantagem da randomização pareada é que ela complica o recrutamento e a randomização, exigindo que um participante elegível aguarde até que um par adequado seja identificado para então poder ser randomizado. Além disso, o pareamento geralmente não é necessário em ensaios clínicos de grande porte nos quais a alocação aleatória equilibra os grupos em relação às variáveis basais. No entanto, uma versão atrativa desse delineamento pode ser usada quando as circunstâncias permitem um contraste sobre os efeitos do tratamento e do controle em duas partes do mesmo indivíduo. No Diabetic Retinopathy Study, por exemplo, cada participante teve um olho aleatoriamente alocado para tratamento com fotocoagulação e o outro serviu como controle (18).

Cegamento

As intervenções devem, sempre que possível, ser planejadas de tal forma que os participantes do estudo, os membros da equipe com os quais eles têm contato, os responsáveis pelas aferições e aqueles que avaliam e adjudicam os desfechos não saibam para qual dos grupos cada um dos participantes foi alocado. Quando não é possível cegar todas essas pessoas, é altamente desejável que o maior número possível seja cegado (por exemplo, cegando sempre os funcionários responsáveis pela aferição dos desfechos). Em um ensaio clínico randomizado, **o cegamento** *é tão importante quanto a* **randomização**. A randomização minimiza a influência de variáveis confundidoras no momento da randomização, mas não tem impacto sobre diferenças que ocorrem entre os grupos durante o seguimento (Tabela 10.2). O cegamento minimiza as fontes de viés pós-randomização, como cointervenções e avaliação e adjudicação enviesadas

dos desfechos.

TABELA 10.2 A randomização elimina o confundimento por variáveis basais, e o cegamento elimina o confundimento por cointervenções

EXPLICAÇÃO PARA A ASSOCIAÇÃO		ESTRATÉGIA PARA ELIMINAR A EXPLICAÇÃO RIVAL
1. Acaso		Igual a estudos observacionais
2. Viés		Igual a estudos observacionais
3. Efeito-causa		(Não é uma explicação possível em estudos experimentais)
4. Confundimento	→ Variáveis confundidoras pré-randomização	Randomização
	→ Variáveis confundidoras pós-randomização (cointervenções)	Cegamento
5. Causa-efeito		

Como já discutido anteriormente no item Escolha do controle, o uso do cegamento para prevenir o viés decorrente de **cointervenções** – medicamentos, terapias ou comportamentos além da intervenção sob estudo e que alteram o risco de desenvolver o desfecho de interesse. O segundo objetivo importante do cegamento é **minimizar vieses na avaliação e na adjudicação dos desfechos**. Em um ensaio não cego, o investigador pode estar inclinado a buscar com maior atenção desfechos no grupo que não recebe tratamento ou mesmo a diagnosticar o desfecho com maior frequência nesse grupo. Por exemplo, em um ensaio não cego sobre estatinas, pode haver uma tendência a perguntar a participantes do grupo experimental sobre dores musculares, levando à solicitação de exames para estabelecer o diagnóstico de miosite. O cegamento dos sujeitos é particularmente importante quando os desfechos forem baseados no autorrelato de sintomas.

Após a identificação de uma possível ocorrência do desfecho, pode ser necessário adjudicá-lo. Por exemplo, quando o desfecho do ensaio clínico é infarto agudo do miocárdio, os investigadores geralmente coletam dados sobre sintomas, achados eletrocardiográficos e enzimas cardíacas. Especialistas cegos em relação ao grupo de tratamento então usam esses dados e definições específicas para adjudicar se o infarto do miocárdio ocorreu ou não. Achados do Canadian Cooperative Multiple Sclerosis Trial ilustram a importância do cegamento para a adjudicação não

enviesada dos desfechos (19). Indivíduos com esclerose múltipla foram aleatoriamente alocados para uma combinação de plasmaferese, ciclofosfamida e prednisona, ou para plasmaferese fictícia e placebo das medicações. No final do estudo, a gravidade da esclerose múltipla foi avaliada com o uso de um exame estruturado por neurologistas cegos para a alocação do tratamento e também por neurologistas não cegos. O tratamento não foi eficaz com base na avaliação dos neurologistas cegados, porém houve eficácia estatisticamente significativa com base na avaliação dos neurologistas não cegados. Os neurologistas não cegos não estavam intencionalmente buscando enviesar o desfecho do ensaio clínico, porém há um forte desejo humano de ver os pacientes melhorarem após o tratamento, especialmente se o tratamento puder causar dor ou for potencialmente danoso. O cegamento minimiza essa adjudicação enviesada do desfecho.

A avaliação cega pode ser menos importante se o desfecho for "duro" (*hard*), como óbito ou medições automatizadas, sobre os quais há pouca possibilidade de avaliação enviesada. A maioria dos outros desfechos, como causa do óbito, diagnóstico de doença, aferições físicas, escalas de questionários e doenças autorrelatadas, são suscetíveis a avaliação e adjudicação enviesadas.

Após a conclusão de um ensaio clínico, pode valer a pena avaliar se os participantes e os investigadores não foram "descegados". Para tanto, eles são convidados a tentar adivinhar as designações de tratamento. Caso uma proporção maior que o esperado acertasse, a discussão do artigo publicado sobre os achados deveria incluir uma avaliação sobre os potenciais vieses desse "descegamento" parcial.

O que fazer quando o cegamento for impossível

Em alguns casos, o cegamento é difícil ou até mesmo impossível, seja por motivos técnicos ou de ordem ética. Por exemplo, é difícil cegar participantes alocados para intervenções educacionais, nutricionais ou de atividade física. Intervenções cirúrgicas muitas vezes não podem ser cegadas, porque pode não ser ético realizar uma cirurgia fictícia no grupo-controle. Contudo, qualquer cirurgia sempre está associada a algum risco, de modo que é muito importante determinar se o procedimento realmente é eficaz. Por exemplo, um ensaio clínico randomizado recente mostrou

que o desbridamento artroscópico da cartilagem do joelho não era mais eficaz do que a artroscopia com falso desbridamento para o alívio da dor de joelho da osteoartrose (20). Nesse caso, o risco aos participantes no grupo-controle pode ser considerado pequeno se o ensaio clínico prevenir que milhares de pacientes sejam submetidos a um procedimento ineficaz.

Se a intervenção não puder ser mascarada, deve-se pelo menos limitar o máximo possível as potenciais cointervenções e assegurar que os indivíduos que farão a avaliação e adjudicação dos desfechos estejam cegados. Por exemplo, um investigador que está testando o efeito da ioga na redução de fogachos da menopausa poderia instruir os participantes de ambos os grupos a evitarem iniciar novos medicamentos, atividades de relaxamento, bem como outros tratamentos para o fogacho até o término do estudo. Além disso, a equipe responsável por coletar informações sobre a gravidade dos fogachos poderia ser diferente daquela que fornece o treinamento de ioga.

RESUMO

1. Um **ensaio clínico randomizado** bem delineado e conduzido pode fornecer a **inferência causal mais definitiva** para orientar diretrizes para a **prática clínica baseada em evidências**.
2. A **escolha da intervenção e da sua dose** são decisões difíceis e que dependem do balanço entre **eficácia** e **segurança**; outras considerações são a **relevância** para a prática clínica, a possibilidade de **cegamento** e a utilização ou não de uma **combinação de medicamentos**.
3. Quando possível, o **grupo de comparação** deveria ser um **controle-placebo** que permite que os participantes, os investigadores e a equipe do estudo sejam **cegados**.
4. **Desfechos clinicamente relevantes**, como dor, qualidade de vida, desenvolvimento de câncer e óbito, são os desfechos mais importantes em ensaios clínicos. **Desfechos intermediários**, como carga viral do HIV, são **marcadores substitutos** válidos para os desfechos clínicos quando as mudanças que o tratamento induz nesses marcadores predisserem mudanças no desfecho clínico.
5. Geralmente é útil **medir mais de uma variável de desfecho**, mas

combiná-las em **desfechos compostos** requer uma avaliação cuidadosa. Um **único desfecho primário** deve ser especificado para testar a hipótese principal.

6. Todos os ensaios clínicos devem incluir medidas que expressem potenciais **efeitos adversos** da intervenção, incluindo tanto **medidas mais focadas** quanto **medidas mais abertas**, usando esse segundo tipo com moderação. Deve-se também implementar procedimentos que assegurem que **eventos adversos graves (EAGs)** sejam prontamente relatados aos comitês de ética em pesquisa (CEPs) e às agências de financiamento.
7. Os critérios de **seleção de participantes para o estudo** devem identificar sujeitos com maior probabilidade de apresentarem **mais benefícios e menos danos** com o tratamento, bem como adequada **adesão aos protocolos de tratamento e seguimento**. Escolher participantes com **alto risco** para o desfecho pode **diminuir o tamanho da amostra**, mas pode tornar o recrutamento mais difícil e reduzir a capacidade de generalização dos achados.
8. Devem-se medir **variáveis basais** com parcimônia para poder **descrever** características dos participantes, **medir os fatores de risco** e os valores basais do desfecho e para permitir o exame das **modificações de efeito**, nas quais a intervenção tem efeitos diferentes em subgrupos diferentes. Deve-se considerar armazenar amostras de soro, material genético, exames de imagem, e assim por diante, na linha de base, para análise posterior.
9. A **randomização**, que minimiza a influência de **variáveis confundidoras** na linha de base, deve ser inviolável; a **randomização pareada** é uma excelente opção quando factível, e em ensaios clínicos de pequeno porte a **randomização em blocos e estratificada** pode reduzir o impacto da distribuição desigual ao acaso de preditores importantes.
10. O **cegamento** da intervenção é **tão importante quanto a randomização**; o cegamento pode controlar para **cointervenções** e para **avaliação** e **adjudicação** enviesadas do desfecho.

REFERÊNCIAS

1. Ettinger B, Black DM, Mitlak BH, et al. Reduction of vertebral fracture risk in

postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. Multiple Outcomes of Raloxifene Evaluation (MORE) investigators. *JAMA* 1999;282:637–645.

2. The Women's Health Initiative Study Group. Design of the women's health initiative clinical trial and observational study. *Control Clin Trials* 1998;19:61–109.
3. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–440.
4. Cummings SR, Karpf DB, Harris F, et al. Improvement in spine bone density and reduction in risk of vertebral fractures during treatment with antiresorptive drugs. *Am J Med* 2002;112:281–289.
5. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998;280:605–613.
6. Cummings SR, San Martin J, McClung MR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med* 2009;361(8):756–765.
7. Mosca L, Barrett-Connor E, Wenger NK, et al. Design and methods of the Raloxifene Use for The Heart (RUTH) Study. *Am J Cardiol* 2001;88:392–395.
8. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. *JAMA* 2002;288:321–333.
9. McClung M, Boonen S, Torring O, et al. Effect of denosumab treatment on the risk of fractures in subgroups of women with postmenopausal osteoporosis. *J Bone Miner Res* 2011;27:211–218.
10. Cummings SR, Black DM, Thompson DE, et al. Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the fracture intervention trial. *JAMA* 1998;280:2077–2082.
11. Fries R, Shariat K, von Wilmowsky H, et al. Sildenafil in the treatment of Raynaud's phenomenon resistant to vasodilatory therapy. *Circulation* 2005;112:2980–2985.
12. Freiman JA, Chalmers TC, Smith H Jr, et al. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978;299:690–694.
13. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984;3:409–420.
14. Friedman LM, Furberg C, DeMets DL. *Fundamentals of clinical trials*, 4th ed. New York: Springer, 2010.
15. Avins AL. Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials. *J Med Ethics* 1998;24:401–408.
16. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer: the women's health initiative randomized controlled dietary

modification trial. *JAMA* 2006;295:629–642.
17. CDP Research Group. The coronary drug project. Initial findings leading to modifications of its research protocol. *JAMA* 1970;214:1303–1313.
18. Diabetic Retinopathy Study Research Group. Preliminary report on effects of photocoagulation therapy. *Am J Ophthalmol* 1976;81:383–396.
19. Noseworthy JH, O'Brien P, Erickson BJ, et al. The Mayo-Clinic Canadian cooperative trial of sulfasalazine in active multiple sclerosis. *Neurology* 1998;51:1342–1352.
20. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81–88.

CAPÍTULO 11

Delineamentos alternativos para o ensaio clínico randomizado e tópicos relacionados à implementação

Deborah Grady, Steven R. Cummings e Stephen B. Hulley

No capítulo anterior, abordamos o delineamento clássico do ensaio clínico randomizado cego em paralelo: como selecionar e mascarar a intervenção e o controle, alocar aleatoriamente as intervenções, escolher os desfechos, lidar com os eventos adversos, selecionar os participantes e medir as variáveis basais e de desfecho.

Neste capítulo, descreveremos **delineamentos alternativos intergrupos randomizados e não randomizados**, bem como **delineamentos intragrupos**, **estudos cruzados** e **estudos-piloto.** Então abordaremos a **condução do ensaio clínico**, incluindo **a adesão à intervenção e ao seguimento**, bem como a **avaliação e adjudicação dos desfechos**. Por fim, concluiremos com uma discussão sobre questões estatísticas, como **monitoramento interino** para a interrupção precoce do ensaio clínico, **análises de intenção de tratar** e **por protocolo** e o uso de **análises de subgrupos** para identificar modificação de efeito.

DELINEAMENTOS RANDOMIZADOS ALTERNATIVOS

Há uma série de variações da estratégia clássica do ensaio clínico randomizado em paralelo que podem ser úteis em determinadas circunstâncias.

Delineamento fatorial

O **delineamento fatorial** busca responder a duas (ou mais) questões de pesquisa em um único ensaio clínico (Figura 11.1). Um exemplo ilustrativo é o Women's Health Study, que tinha como objetivo testar o efeito da aspirina em dose baixa e da vitamina E no risco de eventos cardiovasculares em mulheres saudáveis (1). Os participantes foram aleatoriamente alocados para quatro grupos, e duas hipóteses foram testadas comparando-se duas metades da coorte do estudo. Primeiramente, a taxa de eventos cardiovasculares em mulheres que usaram aspirina foi comparada com aquela em mulheres que receberam placebo da aspirina (independentemente do fato de que metade das mulheres em cada um desses grupos recebeu também vitamina E); então, a taxa de eventos cardiovasculares nas mulheres que receberam vitamina E foi comparada com aquela nas mulheres que receberam placebo da vitamina E (independentemente do fato de que metade das mulheres em cada um desses grupos recebeu aspirina). Dessa forma, os investigadores puderam obter dois ensaios clínicos completos pelo preço de um.

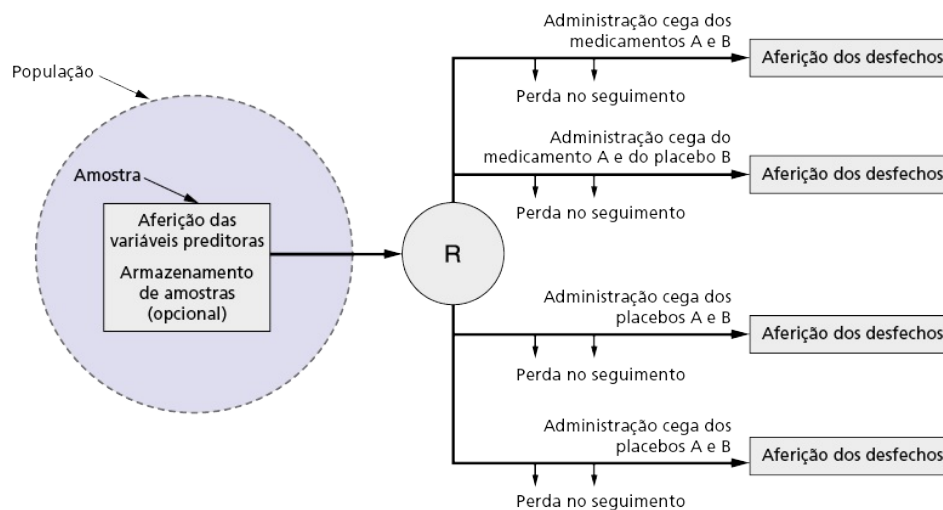



FIGURA 11.1 Para conduzir um ensaio clínico com delineamento fatorial, devem-se seguir os seguintes passos:

- Selecionar uma amostra de participantes a partir de uma população adequada para receber a intervenção.
- Aferir as variáveis preditoras e (se adequado) o nível basal da variável de desfecho.
- Considerar a opção de armazenar soro, imagens, e assim por diante,

para análise posterior.

- Alocar aleatoriamente duas (ou mais) intervenções ativas e seus controles para quatro (ou mais) grupos.
- Seguir as coortes ao longo do tempo, minimizando a perda no seguimento e avaliando a adesão à intervenção e ao controle.
- Aferir as variáveis de desfecho.
- Analisar os resultados, primeiro comparando os dois grupos que receberam a intervenção A (combinados) com os grupos que receberam o placebo A e então comparando os dois grupos que receberam a intervenção B (combinados) com os grupos que receberam o placebo B.

Uma limitação é a possibilidade de **modificação de efeito** (interação): se o efeito da aspirina sobre o risco de doença cardiovascular for diferente em mulheres tratadas com vitamina E em comparação com aquelas não tratadas com vitamina E, há modificação de efeito, e o efeito da aspirina teria de ser calculado separadamente nesses dois grupos. Isso reduziria o poder estatístico dessas comparações, pois apenas metade dos participantes seria incluída em cada análise. De fato, os delineamentos fatoriais podem ser usados para estudar a modificação de efeito, mas ensaios clínicos com esse objetivo são mais complicados e difíceis de implementar, sem contar que é necessário um maior tamanho de amostra e que os resultados podem ser mais difíceis de interpretar. Outras limitações do delineamento fatorial são que a mesma população deve ser adequada para cada uma das intervenções, que a necessidade de múltiplos tratamentos pode interferir no recrutamento e na adesão e que as análises são mais complexas. Isso dito, o delineamento fatorial pode ser muito **eficiente**. Por exemplo, o ensaio clínico do Women's Health Initiative foi capaz de testar o efeito de três intervenções (terapia de reposição hormonal na pós-menopausa, dieta pobre em gorduras e suplementação de cálcio e vitamina D) sobre uma série de desfechos (2).

Randomização por conglomerados

A **randomização por conglomerados** exige que, em vez de alocar indivíduos para intervenções, o investigador aloque aleatoriamente grupos ou conglomerados que ocorrem naturalmente. Um exemplo ilustrativo é um ensaio clínico com jogadores de beisebol de 120 times universitários.

Metade dos times foi alocada a uma intervenção que encorajava que parassem de consumir tabaco de mascar. Observou-se uma redução significativa na taxa de consumo nos times que receberam a intervenção em comparação com os times do grupo-controle (3). Aplicar a intervenção a grupos de indivíduos pode ser mais factível e custo-efetivo do que tratar pacientes individuais. Além disso, adapta-se melhor a questões de pesquisa sobre programas de saúde pública e seus efeitos na população. Isso acontece porque certas intervenções, como dieta com baixos teores de gordura, são difíceis de se implementar em apenas um membro de uma família sem afetar os demais. Quando os participantes em um grupo natural são randomizados individualmente, aqueles que recebem a intervenção provavelmente irão discuti-la com familiares, colegas, membros da equipe ou conhecidos que foram alocados ao grupo-controle.

No delineamento com randomização por conglomerados, as unidades de randomização e análise são grupos, e não indivíduos. Portanto, o tamanho efetivo de amostra é menor do que o número de participantes individuais e o poder é reduzido. O tamanho de amostra efetivo depende da semelhança no efeito da intervenção entre os participantes de cada conglomerado e se situa em algum ponto entre o número de conglomerados e o número total de participantes (4). Outras desvantagens são que a estimativa do tamanho de amostra e a análise dos dados são mais complicadas do que quando é feita randomização individual (4).

Ensaios clínicos com controle ativo: equivalência e não inferioridade

No **ensaio clínico com controle ativo**, o grupo-controle também recebe um tratamento ativo. Esse delineamento é ideal quando já existe um tratamento sabidamente eficaz ou considerado o "tratamento padrão" para uma determinada doença. Esse tipo de ensaio clínico é às vezes denominado **ensaio clínico de eficácia comparativa**, uma vez que se comparam dois tratamentos.

Em alguns casos, o objetivo de um ensaio clínico com controle ativo é mostrar que um novo tratamento é **superior** a um tratamento já estabelecido. Nessa situação, o delineamento e os métodos são semelhantes a um ensaio clínico randomizado controlado por placebo. Na maioria dos casos, no entanto, os investigadores querem estabelecer se um

novo tratamento que tem algumas vantagens sobre um tratamento já estabelecido (p. ex., posologia mais fácil, tratamento menos invasivo ou maior segurança) tem eficácia similar. Nesse caso, é mais adequado realizar um **ensaio clínico de equivalência** ou de **não inferioridade**.

Os **métodos estatísticos** para ensaios clínicos de equivalência e de não inferioridade são diferentes daqueles para ensaios clínicos que buscam mostrar que um tratamento é superior a outro. Em um ensaio clínico planejado para mostrar que um tratamento é superior, a análise-padrão utiliza testes de significância estatística para aceitar ou rejeitar a hipótese nula de que não há diferença entre os grupos. No ensaio clínico que busca mostrar que um novo tratamento é equivalente ao tratamento-padrão, o objetivo ideal seria poder *aceitar* a hipótese nula de que não há diferença. Entretanto, para provar que não há diferença alguma entre tratamentos (nem mesmo uma diferença pequena), seria necessário um tamanho de amostra infinito. Assim, a solução prática é planejar o tamanho de amostra e a análise dos dados com uma abordagem baseada no intervalo de confiança (IC). Para tanto, determina-se onde se situa o IC para o efeito do novo tratamento, comparado ao tratamento-padrão, tendo como referência um delta (“Δ”) pré-especificado, que estabelece uma diferença inaceitável entre os dois tratamentos (5, 6). A equivalência ou não inferioridade é estabelecida, para um nível de confiança especificado pelo IC, quando o IC em torno da diferença na eficácia entre os tratamentos não incluir o Δ (Figura 11.2). Em ensaios clínicos de equivalência, essa é uma questão bicaudal, uma vez que eles buscam mostrar que um novo tratamento não é pior nem melhor que o tratamento-padrão. No entanto, não é comum os investigadores quererem mostrar que um tratamento não *é nem melhor nem pior* que um tratamento estabelecido. O mais comum é querer mostrar que um novo tratamento que apresenta algumas vantagens não é inferior ao tratamento convencional. A natureza unicaudal do delineamento de não inferioridade também tem a vantagem de permitir que o investigador opte por um tamanho de amostra menor ou por um alfa menor; essa última opção geralmente é a preferida (p. ex., 0,025 ao invés de 0,05), para ser conservador.

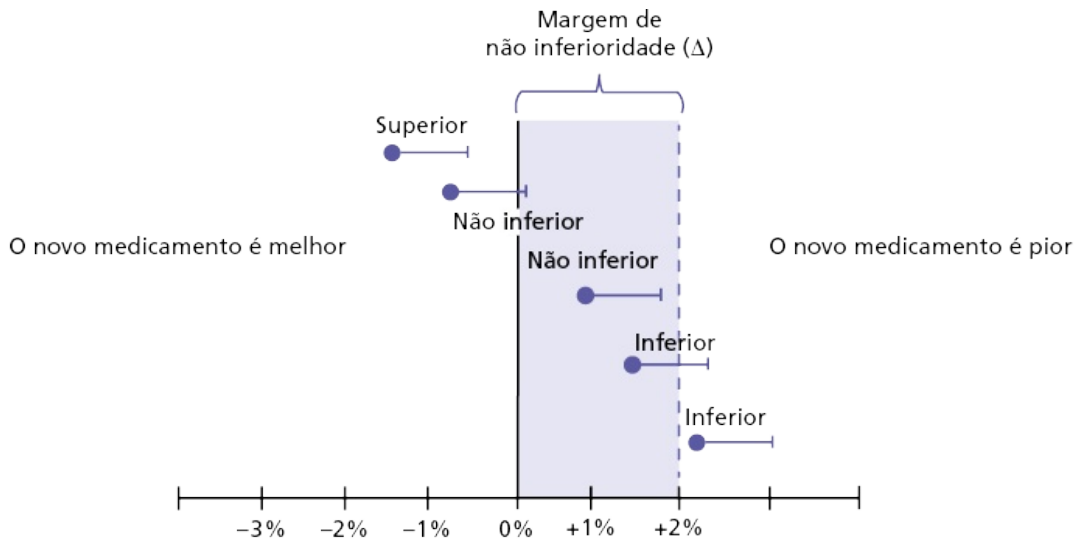



FIGURA 11.2 Possíveis desfechos em um ensaio clínico de não inferioridade comparando um novo medicamento à varfarina para a redução de risco de acidente vascular encefálico em pacientes com fibrilação atrial, com a margem de não inferioridade (delta) estabelecida como sendo +2%. Os intervalos de confiança de 95% unilaterais em torno da diferença na taxa de acidente vascular encefálico entre a varfarina e o medicamento novo são mostrados, ilustrando os desfechos para superioridade, inferioridade e não inferioridade.

Um dos elementos mais difíceis do planejamento de um ensaio clínico de não inferioridade é estabelecer a **margem de não inferioridade** (Δ), isto é, a perda de eficácia com o novo tratamento que seria considerada inaceitável (7). Essa decisão baseia-se em considerações estatísticas e clínicas referentes à potencial eficácia e às possíveis vantagens do novo tratamento, um processo decisório que requer experiência no assunto (8) (ver o Apêndice 11A para um exemplo de como isso é feito). Os ensaios clínicos de não inferioridade geralmente precisam ser maiores do que os ensaios clínicos controlados por placebo, uma vez que a diferença aceitável entre o tratamento novo e o já estabelecido geralmente é menor do que a diferença esperada entre um tratamento novo e o placebo.

É importante ter em mente que a não inferioridade pode não significar que o tratamento estabelecido e o tratamento novo são ambos eficazes – eles podem ser igualmente não eficazes ou mesmo danosos. Para um novo tratamento avaliado em um ensaio clínico de não inferioridade ser considerado mais eficaz do que o placebo, devem haver evidências

prévias sólidas que apoiem a eficácia do tratamento estabelecido. Além disso, o delineamento do ensaio clínico de não inferioridade deveria ser tão similar quanto possível aos ensaios clínicos que estabeleceram a eficácia do tratamento-padrão, incluindo os critérios de seleção, a posologia do tratamento estabelecido, a adesão ao tratamento-padrão, a duração do seguimento, a perda no seguimento e assim por diante (6, 7). Qualquer problema que reduza a eficácia do tratamento-padrão (arrolar participantes que provavelmente não irão se beneficiar, falta de adesão ao tratamento, perda no seguimento) tornará mais provável que o novo tratamento seja mostrado como não inferior, simplesmente porque a eficácia do tratamento-padrão foi diminuída. Um tratamento novo e menos eficaz pode parecer ser não inferior quando, na verdade, os achados representam apenas um estudo mal conduzido.

Em suma, os ensaios clínicos de não inferioridade e os de equivalência são particularmente úteis quando um novo tratamento tem vantagens importantes, como menor custo, posologia mais fácil ou maior segurança. É difícil justificar ensaios clínicos de grande porte para testar um novo medicamento com eficácia semelhante a outros tratamentos já em uso se não houver nenhuma dessas vantagens. É também sempre importante ter em mente que os ensaios clínicos de não inferioridade e de equivalência podem produzir conclusões enganadoras de que dois tratamentos são equivalentes, quando na verdade o ensaio clínico foi mal conduzido.

Delineamentos adaptativos

Os ensaios clínicos são geralmente conduzidos de acordo com um protocolo que não muda durante a realização do estudo. Entretanto, para alguns tipos de tratamentos e doenças, é possível monitorar os resultados do estudo na medida em que ele progride e **mudar o delineamento com base em análises interinas** dos resultados (9). Por exemplo, considere um ensaio clínico sobre diversas dosagens diferentes de um novo tratamento para dispepsia não ulcerosa. O delineamento inicial pode ter planejado arrolar 50 participantes a um grupo placebo e 50 participantes a cada uma de três dosagens diferentes de um tratamento de 12 semanas ao longo de um período de arrolamento de um ano. A revisão dos resultados após os primeiros 10 participantes em cada grupo terem completado quatro semanas de tratamento pode revelar que há uma tendência de alívio

da dispepsia apenas no grupo com a maior dose. Pode ser mais eficiente parar de arrolar participantes para os dois grupos com menor dose e continuar randomizando apenas para o grupo com maior dose e para o placebo. Outros aspectos de um ensaio clínico que poderiam ser mudados com base na análise interina dos resultados incluem aumentar ou diminuir o **tamanho de amostra** ou a **duração** do ensaio clínico se os resultados interinos indicarem que a magnitude de efeito ou a taxa dos desfechos diferem do que se imaginou inicialmente.

Os delineamentos adaptativos são factíveis apenas para tratamentos que produzem desfechos que são medidos e analisados em uma etapa suficientemente precoce do curso do ensaio clínico, permitindo que o delineamento seja alterado nos estágios posteriores do estudo. Para prevenir viés, as regras de como o delineamento pode ser alterado devem ser estabelecidas antes de o estudo iniciar, e as análises interinas e a consideração sobre alterar o delineamento devem ser feitas por um comitê independente de monitoramento dos dados e da segurança, que revisa os dados de forma aberta (não cegada). Realizar múltiplas análises interinas aumenta a probabilidade de encontrar um resultado favorável apenas devido à variação pelo acaso, sendo importante considerar o aumento do risco de erro tipo I ao analisar os resultados.

Além de serem mais complexos de conduzir e analisar, os delineamentos adaptativos requerem que, na obtenção do consentimento informado, seja abordado o escopo das possíveis alterações que poderão ser feitas no delineamento do estudo. Além disso, é difícil estimar o custo de um ensaio clínico adaptativo e os recursos específicos necessários para completá-lo. Apesar dessas precauções e limitações, os delineamentos adaptativos são eficientes e podem ser valiosos, especialmente durante o desenvolvimento de um novo tratamento. Eles podem permitir a identificação mais precoce da melhor dose e duração de tratamento e assegurar que uma proporção elevada de participantes receberá o melhor tratamento.

DELINEAMENTOS NÃO RANDOMIZADOS

Delineamentos não randomizados intergrupos

Em estudos que comparam grupos que não foram randomizados, é muito mais difícil controlar para variáveis confundidoras. Por exemplo, em um estudo sobre os efeitos da cirurgia de revascularização miocárdica em comparação com a angioplastia percutânea, se os clínicos puderem decidir quais pacientes serão submetidos a cada procedimento, em vez de isso ser determinado por meio de alocação aleatória, os pacientes submetidos à cirurgia provavelmente serão diferentes dos pacientes submetidos à angioplastia. Métodos analíticos podem ajustar para fatores basais distribuídos de forma desigual nos dois grupos de estudo, mas essa estratégia não lida com o problema do confundimento não aferido. Ao se comparar os achados de estudos randomizados e não randomizados sobre a mesma questão de pesquisa, os benefícios aparentes da intervenção são frequentemente maiores nos estudos não randomizados, mesmo após o ajuste estatístico para as diferenças nas variáveis basais (10). O problema do confundimento em ensaios clínicos não randomizados pode ser significativo e não é necessariamente eliminado pelo ajuste estatístico (11).

Às vezes, os sujeitos são alocados a grupos de estudo por um mecanismo **pseudoaleatório**. Por exemplo, pode-se alocar ao grupo de tratamento todos os participantes cujo número do prontuário seja um número par. Tais delineamentos podem oferecer vantagens logísticas, mas a previsibilidade da alocação permite que o investigador ou a equipe da pesquisa altere o esquema, manipulando a sequência ou a elegibilidade de novos participantes.

Outras vezes, os sujeitos são alocados a grupos de estudo de acordo com certos critérios clínicos. Por exemplo, pode-se determinar que os pacientes diabéticos recebam insulina quatro vezes por dia ou insulina de longa ação, dependendo de sua disposição em aceitar quatro injeções diárias. O problema é que aqueles dispostos a receber quatro injeções diárias podem diferir daqueles não dispostos (por exemplo, é possível que eles tenham maior adesão a outras orientações sobre sua saúde), e essa pode ser a causa de qualquer diferença nos desfechos observada entre os grupos de tratamento.

Os **delineamentos não randomizados** às vezes são escolhidos devido à falsa crença de que são mais éticos, porque permitem que o sujeito ou seu médico escolham a intervenção. Na verdade, um estudo só será ético se a

qualidade de seu delineamento possibilitar uma resposta correta para a questão de pesquisa. A probabilidade de um resultado conclusivo e correto é maior em estudos randomizados do que em estudos não randomizados. Além disso, a base ética para qualquer ensaio clínico é a incerteza sobre se a intervenção será benéfica ou danosa. Essa incerteza, denominada **equipolência**, significa que a escolha de uma intervenção baseada em evidências ainda não é possível, justificando a alocação aleatória.

Delineamentos intragrupos

Os delineamentos que não incluem um grupo controle separado são úteis para certas questões de pesquisa. Em um **delineamento de séries temporais**, aferições são feitas antes e depois de cada participante receber a intervenção (Figura 11.3). Assim, cada paciente serve como seu próprio controle na hora de avaliar os efeitos do tratamento. Isso significa que características individuais, como sexo, idade e fatores genéticos, não são meramente balanceadas (como ocorre em estudos intergrupos), mas são, na verdade, eliminadas como variáveis confundidoras.

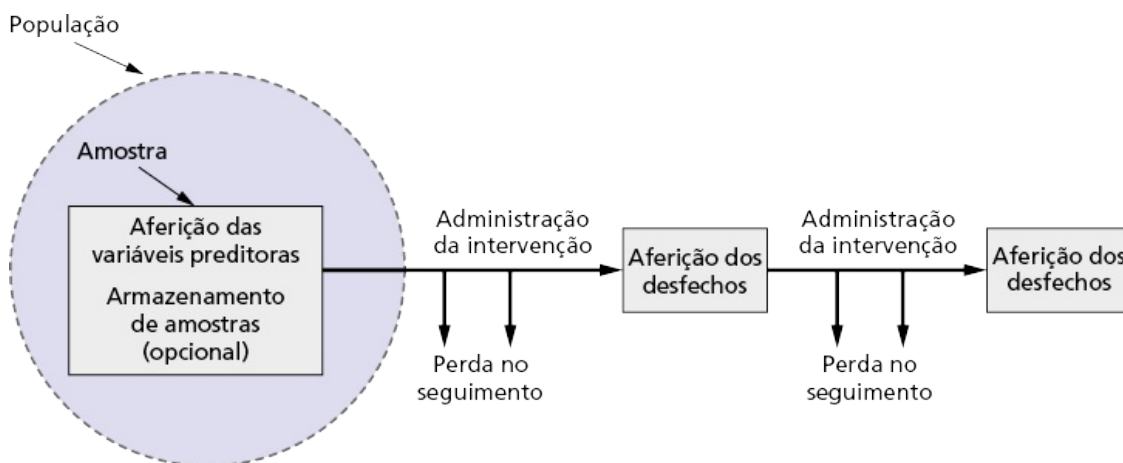



FIGURA 11.3 Para conduzir um ensaio clínico com delineamento de séries temporais, deve-se:

- Selecionar uma amostra de participantes a partir de uma população adequada para receber a intervenção.
- Medir as variáveis preditoras e (se adequado) o nível basal da variável de desfecho.
- Considerar a opção de armazenar soro, imagens, e assim por diante,

para análise posterior.

- Aplicar a intervenção à coorte inteira.
- Seguir a coorte ao longo do tempo, minimizando a perda no seguimento e avaliando a adesão à intervenção.
- Medir as variáveis de desfecho.
- Remover a intervenção, continuar o seguimento e aferir a variável de desfecho novamente e, então, reiniciar a intervenção, e assim por diante.

A principal desvantagem dos delineamentos intragrupos é a falta de um grupo-controle concorrente. Assim, a aparente eficácia da intervenção pode, na verdade, ser causada pelos seguintes fatores: **efeito do aprendizado** (os pacientes têm desempenho melhor em testes cognitivos realizados no seguimento porque aprenderam com o teste realizado na linha de base); **regressão à média** (os participantes selecionados devido à pressão arterial alta na linha de base apresentam pressão arterial mais baixa no seguimento em função apenas da variação aleatória da pressão arterial); e **tendências seculares** (infecções de vias aéreas superiores são menos frequentes no seguimento porque a estação da gripe terminou antes de o seguimento estar concluído). Delineamentos intragrupos podem usar uma estratégia de iniciar e interromper o tratamento repetidas vezes. Se repetidas aplicações e retiradas da intervenção produzem padrões correspondentes no desfecho, isso fornece evidências fortes de que as mudanças se devem ao tratamento. Essa abordagem é útil apenas quando a variável de desfecho responde rápida e reversivelmente à intervenção. Esse tipo de delineamento tem uma aplicação clínica nos **ensaios clínicos "com N igual a um"**, nos quais um paciente pode alternar entre versões ativas e inativas de um medicamento (usando um placebo de aparência idêntica preparado na farmácia local) para detectar sua resposta ao tratamento (12).

Delineamentos cruzados

O **delineamento cruzado** (*crossover*) apresenta características dos ensaios clínicos intra e intergrupos (Figura 11.4). Metade dos participantes é alocada aleatoriamente para iniciar o estudo no período-controle e então mudar para o tratamento ativo; a outra metade faz o

contrário. Essa abordagem permite análises inter e intragrupos. As vantagens são significativas: ela minimiza o potencial de confundimento, pois cada participante serve como seu próprio controle, e a análise pareada aumenta o poder estatístico, sendo necessário um menor número de participantes. No entanto, há também desvantagens importantes, como duplicação da duração do estudo, custos adicionais para medir o desfecho no início e no final de cada período de cruzamento, bem como o aumento da complexidade da análise e da interpretação dos dados, devido ao problema dos **efeitos residuais** (*carryover*). O efeito residual é a influência residual da intervenção no desfecho após sua interrupção. Por exemplo, a pressão arterial pode não voltar aos níveis basais meses após o curso do tratamento com diurético. Para reduzir o efeito residual, pode-se introduzir um período de ***washout*** sem nenhum tratamento, na expectativa de que a variável de desfecho retorne aos níveis basais antes do início da próxima intervenção. No entanto, é difícil saber se o efeito residual foi completamente eliminado. Em geral, os estudos cruzados são uma boa opção quando o número de participantes no estudo for limitado e quando o desfecho responder de forma rápida e reversível a uma intervenção.

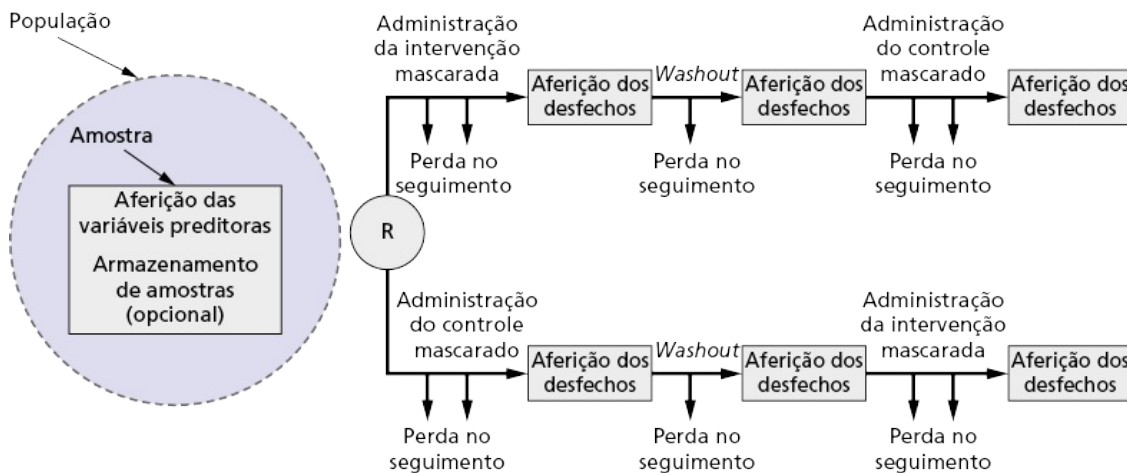



FIGURA 11.4 Em em ensaio clínico randomizado cruzado (*crossover*), devem-se seguir os seguintes passos:

- Selecionar uma amostra de participantes a partir de uma população adequada para receber a intervenção.
- Medir as variáveis preditoras e (se adequado) o nível basal da variável de desfecho.

- Alocar aleatoriamente a intervenção mascarada e o controle.
- Seguir as coortes no tempo, minimizando a perda no seguimento e avaliando a adesão.
- Aferir as variáveis de desfecho.
- Interromper a intervenção e o controle e introduzir um período de *washout* para reduzir o efeito residual, quando adequado.
- Administrar a intervenção ao grupo que anteriormente recebeu o controle e o controle para o grupo que anteriormente recebeu a intervenção e medir os desfechos após seguir as coortes no tempo.

Uma variação do delineamento cruzado pode ser adequada quando a intervenção não puder ser mascarada e ela for vista pelos participantes como muito mais desejável do que o controle, como é o caso de um novo procedimento não invasivo. Nessa situação, pode ser muito difícil encontrar participantes elegíveis que estejam dispostos a serem randomizados; uma forma excelente de lidar com esse problema pode ser randomizar para uma intervenção imediata *versus* um **controle** de **lista de espera** (intervenção postergada). Uma outra situação na qual um **controle** de lista de espera pode ser adequado é quando uma comunidade, escola, governo ou entidade semelhante decide que todos os membros de um grupo devem receber uma intervenção, mesmo havendo evidências limitadas de eficácia. Nesse caso, pode não ser ético randomizar para não receber a intervenção, mas randomizar para uma intervenção postergada pode ser mais aceitável.

O delineamento com fila de espera permite comparar grupos randomizados para intervenção imediata ou lista de espera (grupo-controle). Além disso, os dois períodos de intervenção (imediata e postergada) podem ser combinados para aumentar o poder estatístico da comparação intragrupo (antes e depois da intervenção). Por exemplo, em um ensaio clínico no qual mulheres com miomatose uterina sintomática são randomizadas para um novo tratamento menos invasivo do que a histerectomia (embolização da artéria uterina) *versus* lista de espera, o grupo da lista de espera não receberia tratamento durante o período inicial, mas a embolização da artéria uterina seria oferecida no início do próximo período. No final, podem-se agrupar as medidas de mudanças

intragrupo obtidas no escore de sintomas da miomatose de todos participantes que receberam a intervenção.

Esse delineamento tem a vantagem de tornar o arrolamento muito mais factível em um ensaio clínico onde a intervenção é altamente desejável e de permitir uma comparação randomizada em situações onde todos os participantes elegíveis acabarão recebendo uma intervenção. Entretanto, o desfecho precisa ocorrer após um curto período de tempo (ou o período de lista de espera se tornará proibitivamente longo). Além disso, oferecer a intervenção ao grupo-controle no final do ensaio clínico prolonga a duração do seguimento e pode ter um custo muito elevado.

Ensaios clínicos para a aprovação de novas intervenções

Muitos ensaios clínicos são feitos para testar a efetividade e a segurança de novos tratamentos que poderão ser avaliados pelo Departamento de Alimentos e Medicamentos dos Estados Unidos (Food and Drug Administration, FDA) ou outro órgão regulador. Outro motivo para a realização de ensaios clínicos é averiguar se medicamentos aprovados pelo FDA para um determinado fim poderão ser usados para tratamento ou prevenção de outros problemas de saúde. O delineamento e a condução desses estudos são geralmente os mesmos que os dos demais ensaios clínicos, porém deve-se atentar para as exigências regulatórias.

O FDA publica diretrizes gerais e específicas sobre como esses estudos devem ser conduzidos. (Para mais informações, faça uma busca por “FDA” na internet.) Recomenda-se aos investigadores e ao pessoal envolvido em estudos para a aprovação pelo FDA de um novo medicamento ou dispositivo de uso médico que busquem treinamento específico sobre diretrizes gerais, denominadas **Boas Práticas Clínicas** (*Good Clinical Practices*) (Capítulo 17). Além disso, o FDA fornece diretrizes específicas para estudos sobre certos desfechos. Por exemplo, estudos para obter aprovação do FDA para tratamentos de fogacho em mulheres na menopausa devem incluir participantes com pelo menos sete episódios de fogacho por dia ou 50 por semana. As diretrizes do FDA são periodicamente atualizadas, e diretrizes semelhantes de outras agências reguladoras internacionais também estão disponíveis.

Os ensaios clínicos para a aprovação de novos tratamentos são classificados de acordo com a fase de desenvolvimento em que se situam.

Nesse sistema, os testes ocorrem em uma progressão ordenada, desde experimentos em animais, culturas de células ou tecidos humanos (**etapa pré-clínica**), passando pela administração inicial não cega e não controlada a um número reduzido de voluntários humanos para testar a segurança do tratamento (**fase I**), pequenos ensaios clínicos randomizados ou de séries temporais para avaliar o efeito de diferentes doses sobre os efeitos adversos e sobre biomarcadores ou desfechos clínicos (**fase II**), até ensaios clínicos randomizados de tamanho suficiente para testar a hipótese de que o tratamento melhora a condição-alvo (como pressão arterial) ou reduz o risco de doença (como acidente vascular encefálico) com nível aceitável de segurança (**fase III**) (Tabela 11.1). O FDA geralmente define os desfechos para os ensaios clínicos de fase III que são necessários para obter aprovação para a comercialização de um novo fármaco. A **fase IV** se refere a estudos de maior porte, que podem ser ensaios clínicos randomizados, mas geralmente são estudos observacionais de grande porte conduzidos após a aprovação de um medicamento. Esses estudos muitas vezes são delineados para avaliar a taxa de efeitos colaterais graves quando o medicamento é usado em uma população maior ou para testar usos adicionais do medicamento que poderão ser submetidos à aprovação do FDA. Às vezes, os estudos de fase IV não têm um objetivo científico claro, mas são realizados para introduzir os médicos e os pacientes a novos medicamentos.

TABELA 11.1 Etapas no teste de novas terapias

Etapa pré-clínica	Estudos em cultura de células e em animais
Fase I	Estudos não-cegos e não-controlados em um número reduzido de voluntários para testar a segurança
Fase II	Ensaios clínicos de pequeno porte randomizados ou de séries temporais para testar a tolerância e a diferença de intensidade ou dose da intervenção em biomarcadores ou desfechos clínicos
Fase III	Ensaios clínicos relativamente maiores randomizados, controlados e cegos para testar de forma mais conclusiva o efeito da terapia nos desfechos clínicos e nos eventos adversos
Fase IV	Ensaios clínicos de grande porte ou estudos observacionais conduzidos após a aprovação da terapia pelo FDA para estimar a incidência de efeitos adversos graves incomuns e avaliar outros usos terapêuticos

Estudos-piloto

Para delinear e conduzir um ensaio clínico bem-sucedido é necessário um grande número de dados sobre os seguintes aspectos: tipo, dosagem e duração das intervenções, efeito provável da intervenção sobre o desfecho, potenciais efeitos adversos, factibilidade do recrutamento, da randomização e da manutenção dos participantes no estudo, bem como os prováveis custos. Muitas vezes, a única forma de se obter esses dados é por meio de um bom estudo-piloto.

Os estudos-piloto variam desde um breve teste sobre a factibilidade em um pequeno número de participantes até ensaios clínicos de longa duração em centenas de pacientes, como preparativo para investir em um ensaio clínico multicêntrico de grande porte e com duração de vários anos. Os estudos-piloto devem ser planejados com o mesmo cuidado dispensado ao ensaio clínico principal e precisam ter objetivos e métodos claros. Muitos estudos-piloto concentram-se principalmente na avaliação da **factibilidade,** do **tempo necessário** e do **custo** para recrutar um número adequado de participantes elegíveis e na avaliação de potenciais sujeitos que aceitariam ser randomizados e estariam dispostos a aderir à intervenção. Estudos-piloto também podem ser delineados para demonstrar que as **aferições** planejadas, os **instrumentos** de coleta de dados e os sistemas de **gerenciamento de dados** são factíveis e eficientes. Em estudos-piloto feitos primariamente para testar a factibilidade, um grupo-controle geralmente não é incluído.

Um objetivo importante de muitos estudos-piloto é definir como melhor realizar a **intervenção**, isto é, a frequência, a intensidade e a duração da intervenção que resultam na menor toxicidade e no máximo efeito.

Os estudos-piloto às vezes são usados para fornecer estimativas sobre os parâmetros necessários para estimar o **tamanho da amostra**. Alguns parâmetros cruciais são listados a seguir: estimativas adequadas da taxa do desfecho ou da medida média do desfecho no grupo-placebo, efeito da intervenção sobre o desfecho principal (**magnitude de efeito**) e **variabilidade** estatística do desfecho principal. Na maioria dos casos, é melhor obter essas estimativas a partir de estudos de larga escala que já foram publicados sobre intervenções semelhantes em participantes semelhantes. Na ausência desses dados, usar estimativas de um estudo-piloto pode ser útil, mas os tamanhos de amostra para estudos-piloto geralmente são tão pequenos que as magnitudes de efeito e as variâncias

calculadas são instáveis, com ICs muito amplos.

Muitos ensaios clínicos têm um poder estatístico menor do que o estimado, não porque o efeito da intervenção é menor do que o antecipado, mas porque a **taxa do desfecho** dicotômico no grupo-placebo é muito inferior à esperada. Isso provavelmente ocorre porque os indivíduos que atendem aos critérios de arrolamento para um ensaio clínico e aceitam ser randomizados têm um estado de saúde superior ao da população geral com a doença de interesse. Portanto, é crucial determinar a taxa do desfecho no grupo-placebo, o que pode ser feito avaliando o grupo placebo de ensaios clínicos anteriores com pacientes semelhantes ou randomizando participantes para um grupo-placebo em um estudo-piloto.

Os estudos-piloto devem apresentar um **protocolo curto**, porém **completo** (aprovado pelo Comitê de Ética em Pesquisa da instituição), formulários de coleta de dados e planos de análise. As variáveis incluem as aferições típicas de características da linha de base: dos preditores e dos desfechos que seriam incluídas em um ensaio clínico completo, mas também o número de participantes disponíveis ou acessíveis; o número possível de ser contatado ou que responde usando diferentes técnicas de recrutamento; o número e a proporção elegíveis para o estudo; os elegíveis, mas que se recusam a serem randomizados (ou que afirmam que recusariam); o tempo necessário e o custo do recrutamento e da randomização e estimativas sobre a adesão à intervenção e a outros aspectos do protocolo, incluindo as visitas do estudo. Geralmente é muito útil fazer uma reunião de avaliação com os participantes e com o pessoal envolvido no estudo-piloto após ele ter sido concluído para obter suas opiniões sobre como os métodos do ensaio clínico poderiam ser melhorados.

Um bom estudo-piloto requer tempo substancial e pode apresentar custo elevado, porém ele aumenta consideravelmente as chances de financiamento para ensaios clínicos de grande porte e a probabilidade de que o ensaio clínico seja concluído com sucesso.

CONDUZINDO UM ENSAIO CLÍNICO

Seguimento e adesão ao protocolo

Quando um número significativo de participantes do ensaio clínico não recebe a intervenção, não adere ao protocolo ou é perdido no seguimento, os resultados podem ficar comprometidos por diminuição de poder estatístico ou por viés. Algumas estratégias para **maximizar o seguimento e a adesão** são apresentadas na Tabela 11.2.

TABELA 11.2 Como maximizar o seguimento e a adesão ao protocolo

PRINCÍPIO	EXEMPLO
Escolher sujeitos com maior probabilidade de aderirem à intervenção e ao protocolo	Exigir pelo menos duas visitas antes da randomização
	Excluir os indivíduos que não aderem em um teste inicial anterior à randomização (run-in)
	Excluir aqueles com maior chance de se mudarem ou de não aderirem
Facilitar a intervenção	Usar apenas um comprimido uma vez ao dia se possível
Fazer com que as visitas sejam convenientes e agradáveis	Agendar as visitas com uma frequência que permita manter um contato próximo sem cansar o sujeito
	Marcar as visitas para a noite ou finais de semana, ou coletar as informações por telefone ou e-mail
	Garantir um número suficiente de entrevistadores para que o indivíduo não tenha que esperar
	Fornecer reembolso para deslocamentos
	Manter um bom relacionamento com os sujeitos
Garantir que as medições do estudo não causem dor e sejam úteis e interessantes	Escolher testes não-invasivos e informativos, não disponíveis de outra forma
	Fornecer aos participantes os resultados dos testes de interesse, além de aconselhamento ou encaminhamento para atendimento
Encorajar os sujeitos a permanecer no estudo	Nunca excluir sujeitos do seguimento por causa de violações do protocolo, efeitos adversos ou interrupção da intervenção
	Enviar cartões aos participantes em datas comemorativas
	Enviar boletins periódicos e mensagens via e-mail
	Enfatizar a importância científica da adesão e do seguimento
Localizar os sujeitos perdidos no seguimento	Procurar pessoas que conheçam os sujeitos
	Usar um serviço de localização

O efeito da intervenção (e o poder estatístico do ensaio clínico) se reduz à medida que os participantes não seguem o tratamento. Portanto, é importante escolher um medicamento ou uma intervenção que sejam de fácil aplicação e bem tolerados. Se os participantes tiverem de praticar

uma intervenção comportamental durante horas, sua adesão ficará comprometida. Medicamentos que podem ser tomados em uma única dose diária são mais fáceis de serem lembrados e, portanto, preferíveis. O protocolo deve incluir estratégias que melhorem a adesão, como instruir os participantes a tomar o comprimido sempre na mesma hora, na rotina matinal, fornecendo recipientes que incluam no rótulo o dia da semana, ou enviando lembretes por mensagens no celular.

Também é necessário avaliar a melhor forma de **medir a adesão** à intervenção, usando estratégias como autorrelato, contagem de comprimidos, recipientes para os comprimidos com um *chip* de computador que registre quando o recipiente é aberto e níveis séricos ou urinários de metabólitos. Essas informações permitem identificar os participantes que não estão aderindo adequadamente, de forma que possam ser implementadas estratégias para aumentar a adesão e possibilitando também ao investigador interpretar os achados do estudo de forma apropriada.

Algumas formas de melhorar a **adesão** ao **plano de visitas** e às aferições são discutir com o participante antes da obtenção do consentimento informado sobre o que o estudo envolverá; marcar as visitas para um horário conveniente e com uma equipe com número suficiente de funcionários para evitar demora no atendimento; telefonar ou enviar *e-mail* para o participante no dia anterior à visita; e reembolsar os gastos com deslocamento, estacionamento e outros custos envolvidos.

Perdas no **seguimento** dos participantes do estudo e na aferição do desfecho de interesse podem enviesar os resultados, além de diminuir a credibilidade dos achados e o poder estatístico. Por exemplo, um ensaio clínico sobre o efeito de um *spray* nasal de calcitonina na redução do risco de fraturas osteoporóticas mostrou que o tratamento reduzia o risco de fraturas em 36% (13). Porém, aproximadamente 60% dos indivíduos randomizados foram perdidos no seguimento, e não foi possível determinar se houve fraturas naqueles participantes. Como o número total de fraturas era pequeno, mesmo um número irrisório de fraturas nos participantes perdidos no seguimento poderia ter alterado os achados do estudo, diminuindo, assim, sua credibilidade (14).

Mesmo que os participantes violem o protocolo ou abandonem a intervenção, devem ser mantidos, contabilizando-se os desfechos em

análises de “intenção de tratar” (ver “Analisando os Resultados”, neste capítulo). Muitas vezes os investigadores eliminam do seguimento os participantes que violam o protocolo ao ingressarem em um outro estudo, ao faltarem às consultas ou ao abandonarem a intervenção, e isso pode enviesar os resultados ou torná-los não interpretáveis. Imagine, por exemplo, um medicamento que causa um efeito colateral sintomático que leva o indivíduo a abandonar a intervenção mais frequentemente no grupo de intervenção do que no grupo placebo. Se os participantes que deixaram de usar o medicamento não forem mantidos no seguimento, causando o risco de constituir um viés que pode afetar substancialmente os achados principais, se o efeito colateral estiver associado ao desfecho principal ou a um evento adverso grave (EAG).

As estratégias para garantir um **seguimento** com adesão completa são semelhantes às apresentadas para os estudos de coorte (Capítulo 7). No início do estudo, os participantes devem ser informados sobre a importância do seguimento, e os investigadores devem registrar nome, endereço, e-mail e número telefônico de um ou dois familiares ou pessoas próximas ao participante, que possam sempre informar onde ele se encontra. Além de ajudar o investigador a determinar o estado vital, a possiblidade de contatar os participantes por telefone ou e-mail possibilita acesso a medidas de desfecho substitutas para aqueles sujeitos que se negarem a fazer uma consulta no final do estudo. O ensaio clínico Heart and Estrogen/Progestin Replacement Study (HERS) usou todas essas estratégias: 89% das mulheres retornaram para a consulta final após uma média de quatro anos de seguimento, 8% foram contatadas por telefone para averiguação do desfecho, e o estado vital de cada participante remanescente foi averiguado por meio de cartas registradas, contato com parentes próximos e serviços de localização (15).

O delineamento do ensaio clínico deve facilitar o máximo possível a adesão dos participantes à intervenção e possibilitar que todas as consultas e aferições de seguimento sejam concluídas. Consultas longas e estressantes podem desestimular alguns participantes. É mais provável que eles retornem para consultas que envolvem testes não invasivos, como TC, do que para testes invasivos, como angiografia coronariana. Outra estratégia para aumentar a adesão ao protocolo é coletar as informações do seguimento por telefone ou meios eletrônicos para

aqueles indivíduos com dificuldade de comparecer às consultas. Por outro lado, os participantes podem se desinteressar pelo estudo se ele não propiciar satisfações sociais ou interpessoais. Consultas mensais podem ser cansativas, mas as anuais podem gerar desinteresse. As seguintes estratégias podem ser usadas para maximizar o seguimento, tornando a experiência positiva e agradável para os participantes: planejar as aferições e os procedimentos de forma que não causem dor e sejam interessantes; fazer exames que não estariam disponíveis de outra forma; fornecer os resultados dos exames (a não ser que sejam exames especializados que ainda não estão estabelecidos na prática clínica); enviar boletins, mensagens de texto ou e-mails mostrando consideração; hospedar portais de mídia social; enviar cartões em datas comemorativas; oferecer brindes; e manter um bom relacionamento com a equipe do estudo.

Duas estratégias de delineamento que são específicas aos ensaios clínicos e podem maximizar a adesão e o seguimento são as consultas de triagem anteriores à randomização e um período de teste de entrada (*run-in*). Exigir uma ou duas **consultas de triagem** antes da randomização pode excluir os indivíduos que descobrem que não poderão comparecer às consultas. O segredo é impor obstáculos para a entrada no estudo em nível suficiente para excluir os indivíduos que poderão não aderir ao protocolo, mas ao mesmo tempo cuidando para não excluir aqueles que poderão ter adesão satisfatória.

O período de ***run-in*** (**teste de entrada)** pode ser útil para aumentar a proporção de participantes do estudo que aderem à intervenção e aos procedimentos do seguimento. Durante a linha de base, todos os participantes são incluídos no grupo-placebo. Após um período de tempo especificado (normalmente algumas semanas), apenas aqueles que aderiram à intervenção (p. ex., tomaram pelo menos 80% do placebo) são randomizados. Essa estratégia de excluir os participantes não aderentes antes da randomização pode aumentar o poder do estudo e permitir uma melhor estimativa do espectro total de efeitos da intervenção. Entretanto, o período de *run-in* tem algumas desvantagens: ele retarda a entrada dos participantes no estudo, a proporção que é excluída é geralmente pequena e os participantes randomizados à medicação ativa podem perceber uma mudança na medicação após a randomização, o que leva ao

descegamento. Além disso, ainda não está claro se o período de *run-in* com placebo é mais eficaz para aumentar a adesão do que exigir que os participantes compareçam a uma ou mais consultas de triagem antes da randomização. Na ausência de um motivo específico para suspeitar que a adesão será baixa, provavelmente não é necessário incluir um período de *run-in*.

Uma variante do ***run-in* com placebo** é, em vez de usar placebo, usar medicamento ativo. Além de aumentar a adesão, o ***run-in* com medicamento ativo** permite selecionar participantes que toleram e respondem à intervenção. A ausência de efeitos adversos ou a presença de um efeito desejável do tratamento sobre um biomarcador associado ao desfecho pode ser usado como critério para a randomização. Por exemplo, em um ensaio clínico controlado por placebo sobre o efeito da nitroglicerina sobre a massa óssea, os investigadores usaram um período de *run-in* ativo de uma semana e excluíram as mulheres que interromperam o uso de nitroglicerina devido à cefaleia (16). Essa estratégia maximizou o poder estatístico, pois aumentou a proporção de indivíduos do grupo de intervenção que tolerou o medicamento e provavelmente iriam aderir. No entanto, os achados dos ensaios clínicos que usam essa estratégia podem não ser generalizáveis aos indivíduos excluídos.

É também possível que estudos com *run-in* com tratamento ativo subestimem a taxa de efeitos adversos. Um ensaio clínico sobre o efeito do carvedilol na mortalidade em 1094 pacientes com insuficiência cardíaca congestiva usou um *run-in* com tratamento ativo com duração de duas semanas. Durante esse período, 17 pessoas pioraram da insuficiência cardíaca e 7 morreram (17). Como esses indivíduos não foram randomizados, esses efeitos adversos do tratamento com medicamento ativo não foram computados como desfechos.

Avaliação e adjudicação de desfechos

Os dados para avaliar se um desfecho ocorreu podem vir de inúmeras fontes: autorrelato, questionários padronizados, registros administrativos ou clínicos, exames laboratoriais ou de imagem, aferições especiais, e assim por diante. A maioria dos desfechos autorrelatados, como história de acidente vascular encefálico ou relato de que o participante deixou de

fumar, não são 100% acurados. Portanto, os desfechos autorrelatados de relevância para o estudo devem, quando possível, ser confirmados. É possível adjudicar a ocorrência de uma doença, como acidente vascular encefálico, por meio dos seguintes procedimentos:

1. Elaborar critérios claros para o desfecho (novo *deficit* neurológico persistente com lesão correspondente na TC ou RM);
2. Coletar documentos clínicos necessários para a avaliação (notas de alta hospitalar e laudos radiológicos); e
3. Solicitar que especialistas no assunto revisem cada potencial caso e julguem se os critérios para o diagnóstico foram atendidos.

A adjudicação costuma ser feita por dois especialistas trabalhando de forma independente; eventuais discordâncias são resolvidas entre eles ou por um terceiro especialista. Entretanto, envolver múltiplos especialistas na adjudicação pode ser caro e, para desfechos mais simples em estudos pequenos, ter um único investigador realizando a adjudicação pode ser suficientemente acurado. O importante é que qualquer pessoa envolvida na coleta das informações e na adjudicação dos casos esteja cegada para a alocação do tratamento.

Monitorando ensaios clínicos

Os investigadores devem assegurar que os participantes não serão expostos a uma intervenção danosa, que não lhes seja negada uma intervenção benéfica e que eles não sejam mantidos em um ensaio clínico que provavelmente não irá responder à questão de pesquisa. Cada uma das três considerações a seguir deve ser monitorada durante o ensaio clínico para verificar a necessidade de interromper o estudo precocemente:

- **Interrupção devido a dano:** O primeiro e principal motivo para monitorar ensaios clínicos é garantir que a intervenção não se torne prejudicial. Quando os **danos** estiverem claramente presentes e superarem os benefícios, o estudo deverá ser suspenso.
- **Interrupção devido a benefício:** Se uma intervenção for mais eficaz do que se pensava quando se planejou o estudo, **benefícios** estatisticamente significativos poderão ser percebidos já no início. Se há provas claras

de benefício, pode não ser ético prolongar o estudo e continuar não oferecendo a intervenção aos participantes do grupo-placebo e a outros indivíduos que poderiam se beneficiar.

- **Interrupção devido à futilidade:** Se a probabilidade de responder à questão de pesquisa for muito baixa, pode não ser ético fazer os participantes permanecerem em um estudo que exige tempo e esforço e que pode causar algum desconforto ou risco. Se a duração planejada for de cinco anos e após quatro anos houver pouca diferença nos desfechos dos grupos com tratamento e sem tratamento, o “poder condicional” (probabilidade de rejeitar a hipótese nula no tempo remanescente com base nos resultados disponíveis) torna-se muito pequeno, devendo-se considerar a possibilidade de suspender o estudo. Algumas vezes os ensaios clínicos são interrompidos em um estágio preliminar, porque os investigadores não conseguem recrutar ou manter um número suficiente de participantes para fornecer um poder estatístico adequado para responder à questão de pesquisa, ou porque a adesão à intervenção é muito baixa.

Pode acontecer de a questão de pesquisa ser respondida por outros ensaios clínicos antes do término do estudo em questão. É desejável que mais de um ensaio clínico forneça evidências para uma questão de pesquisa, mas se novas evidências sobre benefício ou dano se tornarem disponíveis durante a realização de um estudo, pode não ser ético continuá-lo.

Os ensaios clínicos devem, em geral, apresentar um **plano de monitoramento interino**. Ensaios clínicos financiados pelos National Institutes of Health (Institutos Nacionais de Saúde dos Estados Unidos, NIH) geralmente requerem monitoramento interino, mesmo que a intervenção seja considerada segura (como intervenção comportamental para perda de peso). É importante determinar, ainda na fase de planejamento, como irá ocorrer esse monitoramento. Em ensaios clínicos de pequeno porte com intervenções provavelmente seguras, os investigadores podem monitorar a segurança ou designar um único monitor independente dos dados e da segurança. Em ensaios clínicos de grande porte e naqueles em que os efeitos adversos da intervenção são desconhecidos ou potencialmente perigosos, o monitoramento interino é

geralmente feito por um comitê (frequentemente denominado Comitê de Monitoramento dos Dados e da Segurança [Data and Safety Monitoring Board, **DSMB**]), composto por especialistas na doença em estudo, bioestatísticos, especialistas em ensaios clínicos, em ética, e, às vezes, um representante do grupo de pacientes sob estudo. Esses especialistas não estão envolvidos no ensaio clínico e não devem ter interesse pessoal ou financeiro na sua continuidade. As diretrizes e os procedimentos desse comitê devem ser detalhados por escrito antes do início do estudo. Orientações para sua elaboração são fornecidas pelo FDA e pelo NIH. Alguns itens que devem ser incluídos nessas diretrizes são listados na Tabela 11.3.

TABELA 11.3 Monitorando um ensaio clínico

Elementos que devem ser monitorados
Recrutamento
Randomização
Adesão à intervenção e ao cegamento
Completude do seguimento
Variáveis importantes
Desfechos
Efeitos adversos
Potenciais co-intervenções
Quem irá monitorar
Os investigadores ou um único monitor independente se o estudo for pequeno e não apresentar maiores riscos
Caso contrário, uma equipe independente de monitoramento dos dados e da segurança
Métodos para o monitoramento interino
Especificar a abordagem estatística e a frequência do monitoramento antecipadamente
Considerar a importância do julgamento e do contexto além das regras estatísticas para a interrupção
Mudanças no protocolo que podem resultar do monitoramento
Término do estudo
Modificações no estudo
Suspender um braço do estudo
Adicionar novas aferições necessárias para monitorar a segurança
Excluir participantes de alto risco
Prolongar a duração do estudo
Aumentar o tamanho de amostra

Interromper um ensaio clínico precocemente deve sempre ser precedido de uma avaliação cuidadosa, contrapondo a responsabilidade ética com os participantes e o avanço do conhecimento científico. Sempre que houver

interrupção precoce, será perdida a chance de se obter resultados mais conclusivos. Trata-se de uma decisão complexa, e os potenciais riscos para os participantes devem ser pesados contra os possíveis benefícios. Os testes de significância estatística que utilizam algum dos métodos que compensam para as múltiplas olhadas nos dados (Apêndice 11B) fornecem informações importantes, porém inconclusivas para o término de um estudo. Deve-se avaliar a consistência de tendências ao longo do tempo, a consistência entre os efeitos observados em desfechos relacionados e ainda o impacto da suspensão precoce do estudo na credibilidade dos achados (Exemplo 11.2).

Há muitos métodos estatísticos para monitorar os resultados interinos de um ensaio clínico. Analisar os resultados repetidas vezes ("múltiplas olhadas") é uma forma de testes múltiplos e aumenta a probabilidade de erro tipo I. Por exemplo, se α = 0,05 for usado para cada teste interino e os resultados forem analisados quatro vezes durante o estudo e novamente no final dele, a probabilidade de um erro tipo I aumentará de 5% para aproximadamente 14% (18). Para resolver esse problema, os métodos estatísticos para monitorar dados interinos normalmente diminuem o *a* para cada teste de forma que no total a se aproxime de 0,05. Há muitas formas de decidir como "gastar o α" (Apêndice 11.B).

Analisando os resultados: intenção de tratar e análise por protocolo

A análise estatística da hipótese principal de um ensaio clínico em geral não apresenta maiores dificuldades. Se o desfecho for dicotômico, o mais simples é comparar as proporções nos grupos usando o teste do **qui-quadrado**. Se o desfecho for contínuo, pode-se usar um **teste *t***, ou, se a distribuição não for normal, uma alternativa não paramétrica. Em muitos ensaios clínicos, a duração do seguimento difere para cada participante, tornando necessário usar métodos de tempo de sobrevida. Modelos estatísticos mais sofisticados, como os dos **azares proporcionais de Cox**, podem alcançar esse fim e, ao mesmo tempo, ajustar para as possíveis distribuições desiguais induzidas pelo acaso nas variáveis confundidoras basais (19).

Um ponto importante que deve ser considerado quando se analisam os resultados de um ensaio clínico é a primazia das abordagens analíticas de

intenção de tratar ao lidar com os indivíduos que foram alocados para o grupo de intervenção e não receberam o tratamento, ou deixaram de usá-lo, e com os indivíduos que foram alocados para o grupo-controle, mas acabaram recebendo o tratamento ativo. Esses dois grupos de indivíduos são denominados ***cross-overs*** (que cruzam grupos). A análise de **intenção de tratar** compara os desfechos nos grupos de estudo, e cada participante é analisado de acordo com sua alocação aleatória, independentemente de ter ou não ter recebido a intervenção que lhe foi designada. Esse tipo de análise pode subestimar o efeito total do tratamento, mas ajuda a proteger contra fontes mais importantes de viés nos resultados.

Uma alternativa à abordagem de intenção de tratar é realizar análises "**por protocolo**", ou seja, que incluem apenas os participantes que aderiram ao protocolo. Isso pode ser definido de diversas formas, mas muitas vezes inclui apenas os participantes em ambos os grupos que aderiram à medicação à qual foram alocados, completaram uma certa proporção das visitas ou aferições e não tiveram outras violações do protocolo. Um tipo especial de análise por protocolo é a análise "**conforme tratado**", na qual são incluídos apenas os participantes que aderiram à intervenção à qual foram alocados. Essas análises *parecem* lógicas, pois os participantes só podem ser afetados por uma intervenção que realmente receberam. No entanto, os participantes que aderem ao estudo e os que não aderem podem diferir em fatores diretamente relacionados ao desfecho. No Postmenopausal Estrogen-Progestin Interventions Trial (PEPI, Ensaio sobre Intervenções com Estrogênio e progestogênio na Pós-Menopausa), 875 mulheres pós-menopáusicas foram aleatoriamente alocadas a quatro diferentes regimes de estrogênio ou estrogênio mais progestogênio e placebo (20). Entre as mulheres alocadas para o braço de estrogênio, 30% haviam abandonado o tratamento após três anos devido à hiperplasia endometrial, precursora do câncer do endométrio. Se essas mulheres fossem eliminadas a partir de uma análise por protocolo, é possível que se deixaria de detectar uma associação entre terapia estrogênica e câncer do endométrio.

A principal desvantagem da análise de intenção de tratar é que os participantes que optarem por não aderir à intervenção planejada também serão incluídos na estimativa sobre os efeitos da intervenção. Dessa forma, se houver um número significativo de abandonos ou *crossovers*

entre os tratamentos, as análises de intenção de tratar subestimarão a magnitude do efeito do tratamento. Por esse motivo, os resultados dos ensaios clínicos muitas vezes são avaliados tanto por intenção de tratar quanto por protocolo. Por exemplo, no ensaio clínico do Women's Health Initiative que avaliou o efeito de estrogênio mais progestogênio sobre o risco de câncer de mama, a razão de azares foi de 1,24 (P = 0,003) na análise de intenção de tratar e de 1,49 na análise conforme tratado (P < 0,001) (21). Se a análise de intenção de tratar e a análise por protocolo produzirem resultados diferentes, os de intenção de tratar geralmente são privilegiados por preservarem o valor da randomização e, ao contrário das análises por protocolo, por enviesarem o efeito estimado apenas na direção conservadora (favorecendo a hipótese nula). No entanto, para estimativas sobre danos (p. ex., os achados sobre câncer de mama), análises conforme tratado ou por protocolo fornecem as estimativas mais conservadoras, pois as intervenções só podem provocar danos em indivíduos expostos.

Os resultados só poderão ser analisados utilizando a abordagem da intenção de tratar se as aferições do seguimento forem completadas para todos os participantes, tenham eles aderido ou não ao tratamento. Portanto, deve-se ter isso sempre como meta.

Análises de subgrupos

Análises de subgrupos são comparações entre grupos randomizados em um subconjunto da coorte do ensaio clínico. O principal motivo para realizar essas análises é identificar **modificação de efeito (“interação”)** nos subgrupos, por exemplo se o efeito de um tratamento difere em homens e mulheres. Essas análises não são bem-vistas por muitos investigadores, pois são facilmente malconduzidas e podem levar a conclusões errôneas. No entanto, se tomadas as devidas precauções, as análises de subgrupos podem fornecer informações complementares e ampliar o leque de inferências que podem ser feitas a partir de um ensaio clínico. Para preservar o valor da randomização, os subgrupos devem ser definidos a partir de medições feitas antes da randomização. Por exemplo, um ensaio clínico sobre a eficácia do denosumabe na prevenção de fraturas mostrou que o medicamento diminuiu o risco de fraturas não vertebrais em 20% em mulheres com baixa densidade óssea. Análises de

subgrupos pré-planejadas revelaram que o tratamento foi eficaz (redução de 35% no risco de fraturas; $P < 0,01$) em mulheres com baixa densidade óssea na linha de base, porém não se mostrou eficaz em mulheres com densidade óssea mais elevada na linha de base ($P = 0,02$ para a modificação de efeito) (22). É importante ressaltar que o valor da randomização é preservado em cada um dos subgrupos: o índice de fraturas em mulheres randomizadas para o tratamento com denosumabe é comparado com o índice de fraturas em mulheres randomizadas para o placebo em cada um dos subgrupos.

Análises de subgrupos baseadas em fatores pós-randomização, como adesão ao tratamento randomizado, não preservam o valor da randomização e frequentemente produzem resultados enganadores. As análises de subgrupos podem produzir resultados desse tipo por diversos motivos. Como os subgrupos são menores do que a população total do estudo, pode não haver poder suficiente para encontrar diferenças importantes. É preciso evitar conclusões do tipo “o tratamento não foi eficaz” em um subgrupo, quando o problema, na verdade, pode ter sido causado pela falta de poder estatístico para encontrar um efeito. Os investigadores muitas vezes examinam resultados em um grande número de subgrupos, aumentando a chance de se encontrar um efeito diferente da intervenção em um subgrupo tão somente pelo acaso. Por exemplo, se 20 subgrupos são examinados, com um $P < 0,05$, pelo mero acaso se poderia esperar a ocorrência uma diferença em um subgrupo. Para abordar esse problema, o planejamento das análises de subgrupos deve ser definido antes do início do estudo e o número de subgrupos analisados deve ser relatado junto com os resultados (23). Ao relatar que as respostas são diferentes nos subgrupos, devem-se incluir evidências de que há interação estatisticamente significativa entre o efeito do tratamento e a característica do subgrupo, devendo a modificação de efeito ser confirmada em um outro estudo antes de ser considerada estabelecida.

RESUMO

1. Há diversas variações do delineamento do ensaio clínico randomizado que podem aumentar substancialmente sua eficiência em determinadas circunstâncias:

a. O **delineamento fatorial** permite dois ou mais ensaios clínicos independentes pelo preço de um.
b. A **randomização por conglomerados** permite estudos eficientes sobre grupos que ocorrem naturalmente.
c. **Ensaios clínicos de não inferioridade ou de equivalência** comparam uma nova intervenção com um “padrão de cuidado” já existente.
d. **Delineamentos adaptativos** aumentam a eficiência ao permitir planejar mudanças baseadas em análises interinas, por exemplo alteração na **dose** do medicamento, no **número** de participantes ou na **duração** do seguimento.

2. Existem também outros delineamentos úteis para o ensaio clínico:
   a. Os **delineamentos de séries temporais** apresentam um único grupo, e os desfechos são comparados em cada participante durante períodos com e sem uma determinada intervenção.
   b. Os **delineamentos cruzados** (*crossover*) combinam os delineamentos intra e intergrupos para aumentar o controle sobre o confundimento (se os **efeitos residuais** [*carryover*] não constituírem um problema) e **minimizar o tamanho da amostra**.
3. Ensaios clínicos para a aprovação de novos medicamentos são classificados como:
   a. **Fase I**, ensaios clínicos de pequeno porte para explorar diferentes opções de dosagem e avaliar a segurança
   b. **Fase II**, ensaios clínicos randomizados de médio porte ou ensaios de séries temporais sobre os efeitos do medicamento em diferentes dosagens
   c. **Fase III**, ensaios clínicos randomizados de grande porte para demonstrar que os benefícios superam os danos, com o objetivo de obter aprovação pelo FDA
   d. **Fase IV**, estudos observacionais de grande porte, após iniciada a comercialização do fármaco, para confirmar os benefícios e detectar efeitos adversos raros
4. **Estudos-piloto** são passos importantes para ajudar a determinar a **aceitabilidade** das intervenções, bem como a **factibilidade**, o **tamanho**, o **custo** e a **duração** dos ensaios clínicos que foram planejados.

5. Ao **conduzir um ensaio clínico**, se um número significativo de participantes **não aderir** à intervenção ou for **perdido no seguimento**, os resultados provavelmente sofrerão de baixo poder estatístico, viés e dificuldade para interpretação.
6. O **monitoramento interino** de dados durante o estudo, por um **comitê independente de monitoramento de dados e da segurança** (*data and safety monitoring board* [DSMB]), é necessário para assegurar a qualidade do estudo e para decidir se o ensaio clínico deve ser **interrompido precocemente** devido a evidências de **dano**, **benefício** ou **futilidade**.
7. As análises de **intenção de tratar** beneficiam-se do controle de confundimento fornecido pela randomização e devem constituir a principal estratégia de análise para avaliar a **eficácia**. As análises **por protocolo**, abordagens secundárias, fornecem uma estimativa da magnitude do efeito nos participantes aderentes (interpretar com cautela), constituindo a análise mais conservadora sobre os efeitos nocivos do tratamento.
8. As **análises de subgrupos** permitem detectar se o efeito do tratamento é modificado por outras variáveis; para minimizar erros de interpretação, o investigador deve especificar os subgrupos antecipadamente, testar a significância estatística das possíveis **modificações de efeito (interações)** e relatar o número de subgrupos examinados.

APÊNDICE 11A

Especificando a margem de não inferioridade em um ensaio clínico de não inferioridade

Uma das decisões mais difíceis ao planejar um **ensaio clínico de não inferioridade** é estabelecer qual nível de perda de eficácia do novo tratamento seria considerado inaceitável (7), referido como "Δ" e denominado **margem de não inferioridade**. Essa decisão baseia-se em considerações estatísticas e clínicas sobre a potencial eficácia e as potenciais vantagens do novo tratamento, além de requerer julgamento por pessoas com experiência no assunto. A seguir apresentamos um exemplo de como isso funciona:

EXEMPLO 11.1 Delineando um estudo para comparar um novo medicamento com a varfarina em pacientes com fibrilação atrial

A varfarina reduz o risco de acidente vascular encefálico em pacientes de alto risco com fibrilação atrial; portanto, um novo medicamento com essa finalidade deveria ser comparado com esse tratamento-padrão já estabelecido. Quando a varfarina é utilizada para reduzir o risco de acidente vascular encefálico em situações como essa, é difícil fazer o ajuste correto da dose, além de serem necessárias múltiplas coletas de sangue para monitorar a anticoagulação, havendo também risco de hemorragia importante. Se houver um novo medicamento que não tenha essas desvantagens, seria lógico preferir esse novo medicamento em relação à varfarina, mesmo que sua eficácia na redução do risco de acidente vascular encefálico seja um pouco inferior.

Uma abordagem para definir o Δ é realizar uma metanálise de ensaios clínicos anteriores sobre varfarina comparada com placebo e definir o Δ como estando em alguma proporção da distância entre a nulidade e o limite inferior do efeito terapêutico da varfarina. Uma vez que os estudos incluídos nas metanálises frequentemente variam muito em termos de qualidade metodológica, uma opção mais

adequada poderia ser basear o Δ nos resultados dos melhores ensaios clínicos, semelhantes em relação aos critérios de entrada, dosagem da varfarina e aferição dos desfechos. É importante definir o Δ de forma que haja grande probabilidade, levando em consideração todos os benefícios e danos, de que o novo tratamento seja superior ao placebo (6, 7).

Suponha que uma metanálise de ensaios clínicos de boa qualidade sobre varfarina comparada com placebo mostre que o tratamento com varfarina reduz a taxa de acidente vascular encefálico em pacientes de alto risco com fibrilação atrial de 10% ao ano para em torno de 5% ao ano (efeito absoluto do tratamento = 5%, IC 95% 4-6%). Tendo em vista as vantagens do novo medicamento, que perda de eficácia seria clinicamente inaceitável? Será que uma eficácia absoluta 2% inferior à da varfarina seria aceitável? Nesse caso o novo tratamento seria considerado não inferior à varfarina se o limite inferior do IC em torno da diferença nas taxas de acidente vascular encefálico entre a varfarina e o novo tratamento for inferior a 2% (Figura 11.2). Em um ensaio clínico de não inferioridade, também é possível que se mostre que o novo tratamento é na verdade superior ao tratamento estabelecido (exemplo no topo na Figura 11.2).

APÊNDICE 11B
Monitoramento interino dos desfechos do ensaio clínico e interrupção precoce

O monitoramento interino dos resultados para decidir sobre interromper ou não um ensaio clínico é uma forma de testes de hipóteses múltiplas, aumentando, assim, a probabilidade do erro tipo I. Para resolver esse problema, o a para cada teste (α_i) é geralmente reduzido de forma que o α global seja 0,05. Há vários métodos estatísticos usados para reduzir o α_i.

Um dos métodos de mais fácil compreensão é o de Bonferroni, onde $\alpha_i = \alpha/N$, sendo N o número total de testes realizados. Por exemplo, se α = 0,05 e cinco testes forem realizados, o α_i para cada teste será 0,01. Esse método tem duas desvantagens: ele usa o mesmo limiar para interromper o estudo em toda e qualquer análise interina e resulta em um a baixo na análise final. A maioria dos investigadores iria preferir usar um limiar mais rigoroso para interromper um ensaio clínico mais precocemente, usando um α próximo de 0,05 para a análise final. Além disso, essa abordagem é muito conservadora, pois pressupõe que cada teste é independente. Análises interinas não são independentes, visto que cada análise sucessiva se baseia em dados cumulativos, alguns dos quais foram incluídos em análises anteriores. Por esses motivos, o método de Bonferroni é pouco usado.

Um método bastante comum sugerido por O'Brien e Fleming (24) é usar um α_i muito baixo para o teste de hipótese inicial e, então, aumentá-lo aos poucos para cada teste sucessivo, de forma que o α_i para o teste final seja próximo ao a global. O'Brien e Fleming apresentam métodos para calcular o α_i quando o investigador escolhe o número de testes a serem realizados e o α global. Em cada teste, $Zi = Z^*(N_i)^{1/2}$, onde Zi = valor Z para o i° teste; Z^* é determinado de forma que alcance o nível global de significância; e N é o número total de testes planejados. Por exemplo, para cinco testes com α global = 0,05 e Z^* = 2,04, o α inicial = 0,00001 e α_5 final = 0,046. É improvável que esse método leve à

interrupção muito precoce de um ensaio clínico, a não ser que haja uma diferença substancial nos desfechos entre os grupos randomizados. Além disso, esse método evita o problema de, no final de um ensaio clínico, a hipótese nula ser aceita quando o valor *P* é 0,04 ou 0,03, mas o α_i para o teste final é diluído para 0,01.

Uma grande desvantagem do método de O'Brien-Fleming é que o número de testes e a proporção de dados a serem testados precisam ser definidos antes do início do estudo. Em alguns ensaios clínicos, testes interinos adicionais são necessários quando ocorrem tendências importantes. DeMets e Lan (25) desenvolveram um método que usa uma função para gasto especificado de a (*specified* α-*spending function*) que fornece limites contínuos para a interrupção. O α_i em um dado momento (ou após uma determinada proporção de desfechos) é determinado por essa função e pelo número de "olhadas" (*looks*) anteriores. Com esse método, não é necessário especificar previamente o número de "olhadas" ou a proporção de dados a serem analisados em cada "olhada". No entanto, obviamente, para cada análise interina conduzida, o α final fica um pouco menor.

Um conjunto diferente de métodos estatísticos baseados em técnicas de *curtailed sampling* permite interromper o ensaio clínico quando os dados futuros provavelmente não irão alterar a conclusão do estudo. O problema dos testes múltiplos torna-se irrelevante, pois a decisão baseia-se apenas na estimativa sobre que dados surgirão no final do estudo. Uma abordagem comum é computar a probabilidade de se rejeitar a hipótese nula no final do ensaio clínico, condicionada nos dados acumulados. Inicialmente, calcula-se o poder condicional pressupondo-se que a *H*o seja verdadeira (i. e., quaisquer desfechos futuros terão distribuição igual nos grupos tratado e controle). Após, pressupõe-se que *H*a seja verdadeira (i. e., os desfechos terão distribuição desigual nos grupos tratado e controle, conforme especificado pela *H*a). Outras estimativas também podem ser usadas para fornecer uma faixa completa de valores plausíveis para a magnitude de efeito. Se o poder condicional para rejeitar a hipótese nula sob qualquer um desses pressupostos for baixo, torna-se pouco provável que a hipótese nula seja rejeitada, e o ensaio clínico poderá ser interrompido.

Exemplos de dois ensaios clínicos que foram interrompidos precocemente são apresentados no Exemplo 11.2.

EXEMPLO 11.2 Dois ensaios clínicos que foram interrompidos precocemente

Cardiac Arrhythmia Suppression Trial (CAST, Ensaio Clínico sobre Supressão de Arritmias Cardíacas) (26). A ocorrência de extrassístoles ventriculares em pessoas que sobreviveram a um infarto agudo do miocárdio (IAM) é um fator de risco para morte súbita. O estudo CAST avaliou o efeito do tratamento antiarrítmico (encainida, flecainida ou moricizina) em pacientes com arritmias ventriculares assintomáticas ou pouco sintomáticas após um IAM sobre o risco de morte súbita. Durante um período médio de 10 meses de seguimento, os participantes tratados com o medicamento ativo tiveram uma mortalidade total mais elevada (7,7% *versus* 3,0%) e uma taxa maior de morte por arritmia (4,5% *versus* 1,5%) do que aqueles alocados para placebo. O estudo foi planejado para durar cinco anos, mas essa diferença grande e com elevada significância estatística levou à sua interrupção após apenas 18 meses.

Physicians' Health Study (Estudo sobre a Saúde de Médicos) (27). O Physicians' Health Study foi um ensaio clínico randomizado sobre o efeito da aspirina (325 mg a cada 2 dias) sobre a mortalidade cardiovascular. O estudo foi interrompido após 4,8 anos, embora o período de seguimento planejado tenha sido de 8 anos. Houve uma redução estatisticamente significativa de infarto não fatal no grupo tratado (risco relativo = 0,56), porém sem diferença no número de óbitos por doença cardiovascular. A taxa de óbitos por doença cardiovascular observada no estudo foi muito inferior à esperada (88 óbitos observados após 4,8 anos de seguimento *versus* os 733 óbitos esperados). O estudo foi interrompido devido ao efeito benéfico da aspirina sobre o risco de infarto não fatal, somado ao poder condicional muito baixo para detectar um impacto favorável sobre a mortalidade cardiovascular.

REFERÊNCIAS

1. Ridker PM, Cook NR, Lee I, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med* 2005;352:1293–1304.
2. The Women’s Health Initiative Study Group. Design of the Women’s Health Initiative clinical trial and observational study. *Control Clin Trials* 1998;19:61–109.
3. Walsh M, Hilton J, Masouredis C, et al. Smokeless tobacco cessation intervention for college athletes: results after 1 year. *Am J Public Health* 1999;89:228–234.
4. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol* 1981;114:906–914.
5. Piaggio G, Elbourne DR, Altman DG, et al. Reporting of non-inferiority and equivalence randomized trials. An extension of the CONSORT Statement. *JAMA* 2006;295:1152–1160.
6. Piaggio G, Elbourne DR, Pocock SJ, et al. Reporting of non-inferiority and equivalence randomized trials. An extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594–2604.
7. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of non-inferiority trials. *Ann Intern Med* 2006;145:62–69.
8. D’Agostino RB Sr., Massaro JM, Sullivan LM, et al. Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Statist Med* 2003;22:169–186.
9. Chang M, Chow S, Pong A. Adaptive design in clinical research: issues, opportunities, and recommendations. *J Biopharm Stat* 2006;16:299–309.
10. Chalmers T, Celano P, Sacks H, et al. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358–1361.
11. Pocock S. Current issues in the design and interpretation of clinical trials. *Br Med J* 1985;296:39–42.
12. Nickles CJ, Mitchall GK, Delmar CB, et al. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. *Pediatrics* 2006;117:2040–2046.
13. Chestnut CH III, Silverman S, Andriano K, et al. A randomized trial of nasal spray salmon calcitonin in postmenopausal women with established osteoporosis: the prevent recurrence of osteoporotic fractures study. *Am J Med* 2000;109:267–276.
14. Cummings SR, Chapurlat R. What PROOF proves about calcitonin and clinical trials. *Am J Med* 2000;109:330–331.
15. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998;280:605–613.
16. Jamal SA, Hamilton CJ, Eastell RJ, Cummings SR. Effect of nitroglycerin ointment on bone density and strength in postmenopausal women. *JAMA* 2011;305:800–805.
17. Pfeffer M, Stevenson L. Beta-adrenergic blockers and survival in heart failure. *N*

Engl J Med 1996;334:1396–1397.
18. Armitage P, McPherson C, Rowe β. Repeated significance tests on accumulating data. *J R Stat Soc* 1969;132A:235–244.
19. Friedman LM, Furberg C, DeMets DL. *Fundamentals of clinical trials*, 3rd ed. St. Louis, MO: Mosby Year Book, 1996.
20. Writing Group for the PEPI Trial. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. *JAMA* 1995;273:199–208.
21. Writing group for WHI investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *JAMA* 2001;288:321-333.
22. McClung MR, Boonen S, Torring O, et al. Effect of denosumab treatment on the risk of fractures in subgroup of women with postmenopausal osteoporosis. *J Bone Mineral Res* 2012;27:211–218.
23. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine—Reporting of subgroup analyses in clinical trials. *NEJM* 2007;357:2189–2194.
24. O’Brien P, Fleming T. A multiple testing procedure for clinical trials. *Biometrics* 1979;35:549–556.
25. DeMets D, Lan G. The alpha spending function approach to interim data analyses. *Cancer Treat Res* 1995;75:1–27.
26. Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989;321:406–412.
27. Physicians’ Health Study Investigations. Findings from the aspirin component of the ongoing Physicians’ Health Study. *N Engl J Med* 1988;318:262–264.

CAPÍTULO 12

Delineando estudos sobre testes médicos

Thomas B. Newman, Warren S. Browner, Steven R. Cummings e Stephen B. Hulley

Os testes médicos, usados, por exemplo, no rastreamento de um fator de risco, no diagnóstico de uma doença e na estimativa do prognóstico de um paciente, são um aspecto importante da pesquisa clínica. Os delineamentos apresentados neste capítulo podem ser usados para **estudar se um determinado teste deve ser realizado e em quem ele deve ser aplicado**.

A maioria dos delineamentos para o estudo de testes médicos lembra os delineamentos observacionais dos Capítulos 7 e 8. Há, no entanto, diferenças importantes entre a maioria dos estudos observacionais e aqueles usados para avaliar testes médicos. A diferença mais importante é que a meta da maioria dos estudos observacionais é identificar associações estatisticamente significativas (Capítulo 5) que representam relações causais. Por sua vez, em estudos sobre testes médicos, não basta demonstrar que um resultado de teste tem uma associação estatisticamente significativa com uma determinada condição para determinar que ele é clinicamente útil, e, além disso, nesses casos, a causalidade geralmente é irrelevante. Portanto, as RC e os valores *P* têm valor secundário nesses estudos, que focam em parâmetros descritivos, como **sensibilidade**, **especificidade** e **razões de verossimilhança**, juntamente com seus **ICs**.

DETERMINANDO SE UM TESTE É ÚTIL

Para um teste ser considerado útil, ele deve sobreviver a uma série de quesitos rigorosos que avaliam sua **reprodutibilidade**, **acurácia** e

factibilidade, e, o que é mais importante, seus **efeitos nas decisões clínicas e nos desfechos** (Tabela 12.1). Respostas favoráveis a esses quesitos são critérios necessários, porém insuficientes, para decidir pela realização de um teste. Por exemplo, se um teste fornecer resultados muito diferentes dependendo de quem o realiza ou de onde ele é feito, dificilmente será útil. Além disso, se um teste raramente acrescentar informações novas, é improvável que influenciará decisões clínicas. Mesmo influenciando decisões clínicas, se elas não se traduzirem em melhora dos desfechos clínicos dos pacientes testados a um risco e custo aceitáveis, o teste também não será útil.

TABELA 12.1 Questões usadas para determinar a utilidade de um teste médico, possíveis delineamentos para respondê-las e estatísticas para relatar os resultados

QUESTÃO	DELINEAMENTOS POSSÍVEIS	ESTATÍSTICAS PARA OS RESULTADOS*
Qual é a reprodutibilidade do teste?	Estudos sobre variabilidade intra e interobservador e intra e interlaboratório	Proporção de concordância, kapa, coeficiente de variação, média e distribuição das diferenças (evitar o coeficiente de correlação)
Qual é a acurácia do teste?	Delineamentos como o transversal, o de caso-controle e o de coorte, nos quais se compara o resultado do teste a um “padrão-ouro”	Sensibilidade, especificidade, valor preditivo positivo e negativo, curvas ROC e razões de verossimilhança (*likelihood ratios*; razões de probabilidades diagnósticas)
Com que frequência os resultados do teste afetam as decisões clínicas?	Estudos sobre rendimento diagnóstico, análises de decisão clínica pré e pós-teste	Proporção de anormais, proporção com resultados discordantes, proporção de testes que levam a mudanças nas decisões clínicas, custo por resultado anormal ou por mudança na decisão clínica
Quais são os custos, os riscos e a aceitabilidade do teste?	Estudos prospectivos ou retrospectivos	Custos médios, proporções com efeitos adversos, proporções de indivíduos dispostos a realizar o teste
A realização do teste melhora o desfecho clínico ou produz efeitos adversos?	Ensaios clínicos randomizados ou estudos de coorte ou de caso-controle nos quais a variável preditora é a aplicação do teste e a de desfecho inclui morbidade, mortalidade ou custos relacionados à doença ou ao seu tratamento	Razões de riscos, razões de chances razões de azares, número necessário tratar, taxas e razões de desfechos desejáveis e indesejáveis

* A maioria das estatísticas desta tabela deve ser acompanhada por seus intervalos de confiança.

É claro que, se a aplicação de um teste melhorar os desfechos dos pacientes testados, será possível inferir respostas favoráveis aos outros quesitos. No entanto, estudos sobre se realizar um teste melhora os desfechos dos pacientes são difíceis de realizar. Assim, os potenciais efeitos de um teste sobre desfechos são geralmente inferidos comparando a acurácia, a segurança ou os custos com aqueles de testes já existentes. Ao desenvolver um novo teste diagnóstico ou prognóstico, pode ser útil avaliar quais aspectos da prática clínica atual necessitam de maior aprimoramento. Por exemplo, os testes usados atualmente são confiáveis? Seus custos são adequados? São seguros e de fácil execução?

Tópicos comuns relacionados a estudos sobre testes médicos

- **Espectro da gravidade da doença e dos resultados de testes.** Como a meta da maioria dos estudos sobre testes médicos é inferir sobre populações por meio de medições em amostras, a validade das inferências depende em grande parte da forma como a amostra foi selecionada. Diz-se que um estudo tem **viés de espectro**, quando o espectro da doença (ou da não doença) na amostra difere daquele esperado nos pacientes para os quais o investigador quer generalizar seus resultados. Em um estágio inicial do desenvolvimento de um teste diagnóstico, pode ser adequado investigar se um teste consegue distinguir sujeitos com doença evidente, em fase tardia, de controles saudáveis. Se ele não conseguir, é possível retornar ao laboratório e buscar modificar o teste ou desenvolver um teste novo. Posteriormente, entretanto, quando a questão de pesquisa abordar a utilidade clínica do teste, o espectro de doença e não doença deve ser representativo dos pacientes para os quais o teste será aplicado. Por exemplo, um teste desenvolvido comparando pacientes com câncer de pâncreas sintomático com controles saudáveis poderia posteriormente ser avaliado em uma amostra mais difícil, porém clinicamente realista, como pacientes consecutivos com dor abdominal e perda de peso não explicados.

  O viés de espectro pode resultar de um espectro inapropriado de resultados de testes ou de um espectro inapropriado de doenças. Por

exemplo, imagine um estudo sobre a concordância interobservador entre as leituras de mamografias por diferentes radiologistas. Se lhes for solicitado classificar os exames como normal ou anormal, sua concordância será muito maior se a maioria dos exames "positivos" que o investigador selecionar para eles examinarem tiver sido selecionada por ser claramente anormal e a maioria dos exames "negativos" tiver sido selecionada por não ter qualquer suspeita de anormalidade.

- **Importância do cegamento (mascaramento).** Muitos estudos sobre testes diagnósticos envolvem julgamentos, como a decisão sobre se o resultado de uma radiografia é anormal ou sobre se um paciente preenche os critérios diagnósticos para uma determinada doença. Sempre que possível, deve-se cegar quem interpreta os resultados dos testes, evitando seu acesso a outras informações sobre o paciente testado. Por exemplo, em um estudo sobre o papel da ultrassonografia no diagnóstico de apendicite, quem realiza a ultrassonografia não deve saber dos achados da história ou do exame físico.[1] Da mesma forma, o patologista que faz o julgamento final sobre quem tem ou não apendicite (o padrão-ouro com o qual os resultados das ultrassonografias serão comparados) não deve saber dos resultados das ultrassonografias. O cegamento previne que vieses, preconceitos e informações provenientes de outras fontes afetem esses julgamentos.

- **Fontes de variação, capacidade de generalização e estratégia de amostragem.** Para determinadas questões de pesquisa, as diferenças entre os pacientes são a principal fonte de variação nos resultados de um teste. Por exemplo, alguns lactentes com bacteriemia terão uma contagem elevada de leucócitos, e outros não. É pouco provável que a proporção de bacteriêmicos com contagem elevada de leucócitos varie muito em função de quem coleta o sangue e de que laboratório o analisa. Por outro lado, os resultados podem depender também da pessoa que aplica os testes ou do ambiente em que são realizados. Por exemplo, a sensibilidade, especificidade e confiabilidade interobservador na interpretação de mamografias depende da habilidade e da experiência do profissional, assim como da qualidade do equipamento. Quando a acurácia ou os custos variam com a pessoa que

interpreta o exame ou com a instituição, devem-se amostrar vários profissionais que realizam a interpretação e várias instituições para avaliar a consistência dos resultados.

- **Padrão-ouro para o diagnóstico.** Algumas doenças têm um **padrão-ouro** que é geralmente aceito como indicativo da presença (ou ausência) da doença, como, por exemplo, o exame histopatológico para a identificação de câncer. Para outras doenças, o **padrão-ouro** é uma "definição", como a de que para haver doença coronariana é necessária uma obstrução de 50% ou mais em pelo menos uma artéria coronária principal na angiografia. Para outras doenças, como as reumáticas, o paciente precisa apresentar um número mínimo de sinais, sintomas ou anormalidades específicas em exames laboratoriais. Obviamente, se quaisquer sinais, sintomas ou exames laboratoriais usados para diagnosticar uma doença fizerem parte do padrão-ouro, um estudo que comparasse essas alterações com o padrão-ouro poderia fazê-las parecer falsamente boas. Isso é denominado **viés de incorporação**, uma vez que o teste que está sendo estudado é *incorporado* ao padrão-ouro. Evitar esse viés é um dos motivos previamente mencionados para o cegamento.

  Também é importante considerar se o padrão-ouro é realmente perfeito. Se o padrão-ouro for imperfeito, ele pode levar o teste a parecer pior do que realmente é (quando na verdade o teste é superior ao padrão-ouro), ou ainda melhor do que ele realmente é (se o teste que está sendo avaliado tiver as mesmas deficiências que o padrão-ouro).

- **Em que consiste um resultado de teste positivo?** Especialmente se o teste tiver resultados em escala contínua (como níveis de eritropoietina sérica), o investigador pode sentir-se tentado a examinar todos os resultados nas pessoas com o desfecho (p. ex., anemia da doença crônica) e naquelas sem o desfecho (outros tipos de anemia) e então selecionar o melhor **ponto de corte** para definir um teste como positivo. No entanto, isso é um tipo de **sobreajuste** (isto é, variação aleatória na amostra específica sob estudo de que o desempenho no teste parece melhor do que realmente é na população). Algumas estratégias

melhores são basear o ponto de corte no conhecimento clínico ou biológico a partir de outros estudos ou dividir testes contínuos em intervalos e então calcular as razões de verossimilhança para cada intervalo (ver texto a seguir). Para minimizar o sobreajuste, os pontos de corte que definem os intervalos devem ser especificados *a priori*, ou devem ser utilizados números inteiros que façam sentido clinicamente. O sobreajuste é um problema especialmente em estudos sobre regras de predição clínica, descritos mais adiante neste capítulo.

ESTUDOS SOBRE A REPRODUTIBILIDADE DE TESTES

Às vezes, os resultados dos testes variam dependendo de quando ou de onde eles foram feitos, ou até mesmo de quem os aplicou. A variabilidade **intraobservador** descreve a falta de reprodutibilidade dos resultados quando um mesmo observador ou laboratório realiza o teste repetidas vezes na mesma amostra. Por exemplo, se um radiologista examinar duas vezes a mesma radiografia de tórax, qual a percentagem dos exames em que ele concordará com a sua própria interpretação, pressupondo que não esteja ciente de sua interpretação anterior? A variabilidade **interobservador** descreve a falta de reprodutibilidade entre dois ou mais observadores: se outro radiologista examinar o mesmo filme, em que proporção dos exames ele irá concordar com a interpretação do primeiro radiologista?

Muitas vezes, a questão principal é o nível de reprodutibilidade de um determinado teste (ou mesmo a falta dela). Em outros, a reprodutibilidade é estudada com o intuito de melhorar a qualidade, seja da prática clínica ou de um estudo. Quando a reprodutibilidade for baixa – em função de uma elevada variabilidade intra ou interobservador –, o teste diagnóstico provavelmente não terá utilidade clínica, necessitando ser melhorado ou abandonado.

Os estudos sobre reprodutibilidade avaliam a precisão e não a acurácia ou validade (Capítulo 4), de modo que os observadores podem concordar uns com os outros e ainda assim estarem errados. Quando um padrão-ouro estiver disponível, os pesquisadores que estão avaliando a reprodutibilidade intra e interobservador podem comparar as observações dos sujeitos com um padrão-ouro para determinar a acurácia. Quando não

houver padrão-ouro disponível, os investigadores devem se apoiar em outros métodos para avaliar a validade, descritos no Capítulo 4.

Delineamentos

O delineamento básico para avaliar a reprodutibilidade de um teste baseia-se na comparação dos resultados obtidos por mais de um observador ou em mais de uma ocasião. Para testes diagnósticos realizados em várias etapas, nos quais diferenças em uma única etapa podem afetar a reprodutibilidade, é importante decidir a abrangência do que o estudo irá abordar. Por exemplo, a demonstração de que vários patologistas que trabalham em um mesmo hospital concordam na interpretação de lâminas de citologia cervical pode superestimar a reprodutibilidade global do teste, visto que não foram levadas em consideração a variabilidade no modo de obtenção da amostra e a variabilidade no preparo da lâmina.

Até que ponto é importante isolar cada uma das etapas que podem levar à falta de concordância interobservador depende das metas do estudo. Em geral, os estudos deveriam estimar a reprodutibilidade de todos os procedimentos envolvidos na realização do teste, pois é isso que determina se vale a pena ou não realizá-lo. No entanto, se o investigador estiver desenvolvendo ou aprimorando um teste, ele pode preferir focar em etapas específicas que são mais problemáticas, com o objetivo de aprimorar o processo. Em qualquer uma dessas situações, é importante explicitar no manual de operações o processo exato de obtenção do resultado do teste (Capítulos 4 e 17) e, ao relatar os resultados, descrever esse processo na seção de métodos.

Análise

- **Variáveis categóricas.** A medida mais simples de concordância interobservador é o percentual de observações nas quais a concordância é total. No entanto, quando as observações não estiverem distribuídas uniformemente entre as categorias (p. ex., quando a proporção de “anormais” definidos por um teste dicotômico for muito diferente de 50%), pode ser difícil interpretar o percentual de concordância, uma vez que ela não leva em consideração a concordância que resultaria do fato

de ambos os observadores terem algum conhecimento sobre a prevalência da anormalidade. Por exemplo, se 95% dos sujeitos forem normais, dois observadores que escolherem aleatoriamente quais 5% dos testes eles irão considerar anormais irão concordar que os resultados são “normais” 90% das vezes. O percentual de concordância também não é uma medida ideal quando um teste tem mais de dois resultados possíveis e esses resultados possuem uma ordem intrínseca (p. ex., normal, limítrofe e anormal), uma vez que ele considera a concordância parcial (p. ex., normal/limítrofe) da mesma forma como a ausência total de concordância (normal/anormal). Uma medida melhor de concordância interobservador, denominada **kapa** (κ) (Apêndice 12A), mede o grau de concordância além do que seria esperado a partir do conhecimento dos observadores sobre a prevalência da anormalidade[2] e permite valorizar a concordância parcial. O kapa varia de −1 (ausência total de concordância) a 1 (concordância total). Um kapa de 0 indica que a concordância não foi superior à que seria esperada pelas estimativas dos observadores sobre a prevalência de cada nível de anormalidade. Valores de kapa superiores a 0,8 são geralmente considerados muito bons; níveis entre 0,6 e 0,8 são considerados bons.

- **Variáveis contínuas.** A escolha das medidas de variabilidade interobservador para variáveis contínuas depende do delineamento empregado. Certos estudos medem a concordância entre apenas dois aparelhos ou métodos (p. ex., temperaturas obtidas a partir de dois termômetros). A melhor forma de descrever dados de um estudo como esse é reunir os dados sobre pares de medidas (cada par consistindo em duas medições feitas em momentos próximos em um mesmo sujeito) e relatar a diferença média entre essas medidas pareadas, bem como alguma medida da dispersão dos valores, como o desvio-padrão ou a frequência com que a diferença excede um limiar clinicamente relevante. Por exemplo, se 0,3°C for considerada uma diferença de temperatura corporal clinicamente relevante, um estudo que compara as temperaturas dos termômetros timpânico e retal poderá estimar a diferença média entre as duas medições (± o desvio-padrão) e com que frequência elas irão diferir em mais de 0,3° C.[3]

Outros estudos abordam a variabilidade interensaio, interobservador ou interinstrumento de testes realizados por um grupo grande de técnicos, laboratórios ou aparelhos diferentes. Esses resultados são geralmente sumarizados usando o **coeficiente de variação (CV)**, que é o desvio-padrão de todos resultados obtidos de uma única amostra dividido pela sua média. Muitas vezes, comparam-se os CVs de dois ou mais diferentes ensaios ou instrumentos; aquele com o menor CV é o mais preciso (embora possa não ser aquele com a maior acurácia).

ESTUDOS SOBRE A ACURÁCIA DE TESTES

Todos os estudos nesta seção abordam a questão "Até que ponto o teste fornece a resposta correta?". Isso pressupõe, é claro, que exista um **padrão-ouro** para revelar qual é a resposta correta.

Delineamentos

- **Amostragem.** Estudos sobre a acurácia de testes diagnósticos podem ter delineamentos análogos aos estudos de caso-controle ou transversais. Nos delineamentos de **caso-controle** para a avaliação de testes diagnósticos, indivíduos com a doença e aqueles sem a doença são amostrados separadamente, e os resultados dos testes nos dois grupos são comparados. Como mencionado, a amostragem de casos e controles pode ser apropriada em um estágio inicial no desenvolvimento de um teste diagnóstico, quando a questão de pesquisa é se o teste merece estudos mais aprofundados. Em uma etapa posterior, quando a questão de pesquisa é sobre a utilidade clínica do teste, os espectros de doença e não doença devem assemelhar-se àqueles das pessoas para as quais o teste será aplicado clinicamente; é muito mais difícil responder a essa questão com a amostragem de casos e controles do que com amostras que sejam representativas de toda a população-alvo.

  Estudos sobre testes que fazem a amostragem das pessoas com e sem a doença separadamente estão sujeitos a viés na aferição ou no relato do resultado do teste, uma vez que sua aferição ocorre necessariamente após já haver sido determinado se a pessoa tem ou não a doença. Além

disso, estudos com esse esquema de amostragem geralmente não podem ser utilizados para estimar os **valores preditivos** (como discutido no texto a seguir).

Uma amostra **consecutiva** de pacientes que estão sendo avaliados para um determinado diagnóstico geralmente produz resultados mais válidos e mais facilmente interpretáveis, incluindo os valores preditivos. Por exemplo, Tokuda e colaboradores (3) demonstraram que a intensidade dos calafrios (p. ex., apenas sentir frio *versus* sentir o corpo inteiro tremendo embaixo de um cobertor) era um forte preditor de bacteriemia em uma série de 526 pacientes consecutivos com febre, em um serviço de emergência. Como o arrolamento ocorreu antes de se saber se tinham bacteriemia, o espectro de pacientes representou adequadamente os indivíduos que procuraram serviços de emergência por febre.

Uma variante da amostragem transversal, denominada **teste em tandem**, pode ser usada para comparar dois testes que se imagina serem imperfeitos. Ambos os testes são aplicados em uma amostra representativa de sujeitos, e o padrão-ouro é aplicado naqueles com resultado positivo em pelo menos um dos testes. O padrão-ouro deve ser aplicado também a pelo menos uma amostra aleatória de pacientes com resultados negativos concordantes, para assegurar-se de que eles realmente não têm a doença. Esse delineamento, que permite determinar qual dos testes é mais acurado, sem precisar aplicar um padrão-ouro a todos os sujeitos com resultados negativos, foi usado em estudos que comparavam métodos diferentes de citologia cervical (4).

Para realizar estudos sobre testes prognósticos são necessários **delineamentos de coorte**. No delineamento prospectivo, o teste é aplicado na linha de base, e os sujeitos são seguidos no tempo para verificar quais desenvolvem o desfecho de interesse. O estudo de coorte retrospectivo pode ser usado quando surge um novo teste, como carga viral em pacientes HIV positivos, se houver uma coorte definida previamente disponível com amostras de sangue armazenadas. Nesse caso, a carga viral pode ser medida no soro armazenado para verificar se ela prediz o prognóstico. O delineamento tipo caso-controle aninhado (Capítulo 8) é especialmente atraente quando o desfecho de interesse é raro e o teste apresenta alto custo.

- **Variável preditora: o resultado do teste.** Embora seja mais simples pensar sobre os resultados de um teste diagnóstico como positivos ou negativos, muitos testes têm resultados categóricos, ordinais ou contínuos. Para tirar vantagem de todas as informações que o teste oferece, deve-se geralmente relatar os resultados na forma ordinal ou contínua, em vez de dicotomizar em “normal ou anormal”. A maioria dos testes é mais indicativa de doença se o resultado for muito anormal do que se for apenas levemente anormal, e a maioria possui uma faixa de valores na qual não fornecem muita informação.

- **Variável de desfecho: a doença (ou seu desfecho).** A variável de desfecho em um **estudo sobre um teste diagnóstico** é a presença ou ausência da doença, mais bem determinada a partir de um padrão-ouro. Sempre que possível, deve-se garantir que a avaliação do desfecho não seja influenciada pelos resultados do teste diagnóstico que está sendo estudado. A melhor forma de se fazer isso é cegando aqueles que irão aplicar o padrão-ouro de forma que não saibam os resultados do teste que está sendo avaliado.

  Às vezes, especialmente para testes de rastreamento, a aplicação uniforme do padrão-ouro não é ética ou factível. Por exemplo, Smith-Bindman e colaboradores (5) estudaram a acurácia da mamografia de acordo com características do radiologista que fazia a interpretação. Recomendou-se às mulheres com resultado positivo na mamografia que fizessem testes adicionais, eventualmente chegando ao “padrão-ouro” da avaliação pelo patologista. No entanto, não parecia sensato exigir biópsia de mulheres com mamografia negativa. Então, para avaliar se essas mulheres tinham falso-negativos na mamografia, os autores fizeram *linkagem* dos resultados das mamografias com registros locais de câncer e consideraram a ocorrência ou não do diagnóstico de câncer de mama no ano subsequente como padrão-ouro. Essa solução pressupõe que todos os casos de câncer de mama existentes no momento da mamografia seriam diagnosticados no período de um ano, e que todos os tumores diagnosticados no período de um ano existiam no momento da mamografia. Medir o padrão-ouro de forma diferente dependendo do resultado do teste introduz um potencial de viés,

discutido em maior detalhe no final deste capítulo, mas às vezes essa é a única opção factível.

A variável de desfecho em **estudos sobre testes prognósticos** é aquilo que irá ocorrer com os pacientes que têm uma determinada doença; por exemplo, quanto tempo ainda irão viver, que complicações irão desenvolver ou de que tratamentos adicionais irão necessitar. Novamente, o cegamento é importante, em especial se os médicos que estão atendendo os pacientes tomarem decisões com base nos fatores prognósticos em estudo. Por exemplo, Rocker e colaboradores (6) demonstraram que a estimativa feita pelo médico sobre o prognóstico de um paciente, e não a estimativa da enfermeira, estava associada independentemente à mortalidade na unidade de tratamento intensivo. Uma explicação para isso poderia ser que os médicos fossem mais capazes de estimar a gravidade da doença, mas também poderia ser que a avaliação médica sobre o prognóstico tivesse um efeito maior do que a das enfermeiras sobre a decisão de suspender ou não o suporte de vida. Para distinguir entre essas possibilidades, seria útil obter estimativas sobre o prognóstico atribuído por médicos que não fossem aqueles envolvidos na tomada de decisão sobre suspender ou não o suporte.

Análise

- **Sensibilidade, especificidade e valores preditivos positivo e negativo.**

Quando os resultados de um teste dicotômico são comparados com um padrão-ouro, os resultados podem ser sumarizados em uma tabela 2 × 2 (Tabela 12.2). A **sensibilidade** de um teste é definida como a proporção de sujeitos com a doença para os quais o teste fornece a resposta correta (teste positivo); a **especificidade** é a proporção de sujeitos sem a doença para os quais o teste fornece a resposta correta (teste negativo). Se a amostra de pacientes estudados for representativa do grupo de pacientes para os quais o teste seria utilizado, dois parâmetros adicionais podem ser calculados. O **valor preditivo positivo** é a proporção de sujeitos com resultado positivo e que têm a doença; o **valor preditivo negativo** é a proporção de sujeitos com resultado negativo e que não têm a doença.

TABELA 12.2 Sumarizando os resultados de um estudo sobre um teste dicotômico em uma tabela 2 × 2

		PADRÃO-OURO			
		COM A DOENÇA	**SEM A DOENÇA**	**TOTAL**	
TESTE	***Positivo***	a Verdadeiro-positivo	b Falso-positivo	a + b	Valor preditivo positivo = a/(a + b)
	Negativo	c Falso-negativo	d Verdadeiro-negativo	c + d	Valor preditivo negativo = d/(c + d)
	Total	a + c Sensibilidade = a/(a + c)	b + d Especificidade = d/(b + d)		

Os valores preditivos positivo e negativo podem ser calculados a partir de uma tabela 2 × 2 usando essa fórmula apenas quando a prevalência da doença é (a + c)/(a + b + c + d). Esse não será o caso se os sujeitos com e sem a doença forem amostrados de forma separada (p. ex., 100 sujeitos em cada grupo em um estudo com amostragem de caso-controle).

- **Curvas ROC.** Muitos testes diagnósticos produzem resultados ordinais ou contínuos. Para esses testes, vários valores de sensibilidade e especificidade são possíveis, dependendo do ponto de corte escolhido para definir um teste como positivo. Esse balanço entre sensibilidade e especificidade pode ser mostrado usando-se um método gráfico desenvolvido originalmente na eletrônica, a **curva ROC** (*receiver operator characteristic* – característica operatória do receptor). O investigador seleciona vários pontos de corte e determina a sensibilidade e especificidade em cada ponto. Então, desenha um gráfico colocando a sensibilidade (índice de verdadeiro-positivos) no eixo Y como função de 1 − especificidade (índice de falso-positivos) no eixo X. O teste ideal é aquele que atinge o canto superior esquerdo do gráfico (100% de verdadeiro-positivos e nenhum falso-positivo). Um teste inútil segue a diagonal que vai do canto inferior esquerdo até o canto superior direito, na qual, em qualquer ponto de corte, o índice de verdadeiro-positivos é igual ao índice de falso-positivos (Figura 12.1). A área sob a curva ROC, que varia de 0,5 para um teste inútil a 1,0 para um teste perfeito, resume bem a acurácia global de um teste e pode ser usada para comparar a acurácia de dois ou mais testes.

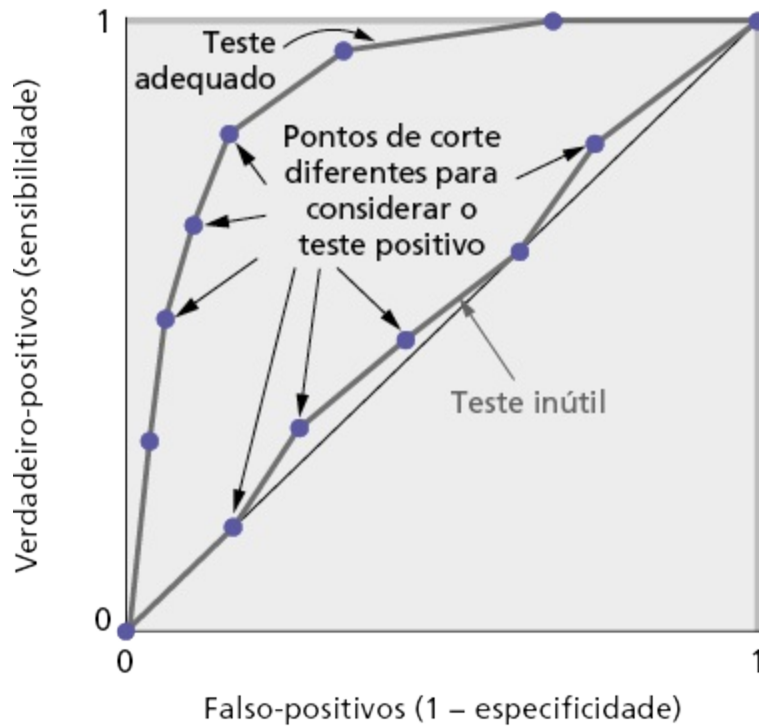



FIGURA 12.1 Curvas ROC para testes bons e para testes inúteis.

- **Razões de verossimilhança (razões de probabilidades diagnósticas ou likelihood ratios).** Embora as informações sobre um teste diagnóstico com resultados contínuos ou ordinais possam ser resumidas por meio da sensibilidade e da especificidade, ou por meio de uma curva ROC, existe uma forma melhor para fazer isso. Razões de verossimilhança ou razões de probabilidades diagnósticas (*likelihood ratios*) permitem que sejam aproveitadas todas as informações disponíveis em um teste. Para cada resultado do teste, a razão de verossimilhança é o quociente entre a probabilidade de se encontrar o resultado em alguém com a doença e a probabilidade de se encontrá-lo em alguém sem a doença.

$$\text{Razão de verossimilhança} = \frac{\text{P (Resultado | Doença)}}{\text{P (Resultado | Ausência de Doença)}}$$

P é lido como "probabilidade de", e o símbolo "|" é lido como "dado (a)". Assim, *P* (Resultado|Doença) é a probabilidade de um resultado dada a doença, e *P* (Resultado|Ausência de Doença) é a probabilidade do resultado dada a ausência de doença. A razão de verossimilhança é

uma razão entre essas duas probabilidades.[4]

Quanto maior a razão de verossimilhança, melhor o resultado do teste para confirmar o diagnóstico de uma doença; uma razão de verossimilhança superior a 100 é muito alta (e rara entre os testes). Por outro lado, quanto menor a razão de verossimilhança (quanto mais próxima ela for de 0), melhor o resultado do teste para descartar a possibilidade de doença. Uma razão de verossimilhança de 1 não fornece nenhuma informação sobre a probabilidade da doença; quando ela é próxima de 1 (digamos, algo entre 0,8 e 1,25), fornece pouca informação relevante.

Um exemplo do uso de razões de verossimilhança é mostrado na Tabela 12.3, que apresenta os resultados dos hemogramas de recém-nascidos em risco para infecções graves (7). Uma contagem de leucócitos inferior a 5.000 células/µL era muito mais comum em lactentes com infecções graves do que em outras crianças. O cálculo da razão de verossimilhança simplesmente quantifica o seguinte: 19% das crianças com infecções graves tinham contagens de leucócitos inferiores a 5.000 células/µL, comparado a apenas 0,52% daquelas sem infecções graves. Portanto, a razão de verossimilhança é 19%/0,52% = 36.

TABELA 12.3 Exemplo de cálculo de razões de verossimilhança para um estudo sobre o papel do hemograma na predição de infecções graves em recém-nascidos (7)

CONTAGEM DE LEUCÓCITOS (POR ML)	INFECÇÃO GRAVE		RAZÃO DE VEROSSIMILHANÇA
	SIM	NÃO	
< 5.000	46	347	
	19%	0,52%	36
5.000-9.999	53	5.103	
	22%	7,6%	2,9
10.000-14.999	53	16.941	
	22%	25%	0,86
15.000-19.999	45	21.168	
	18%	31%	0,58
≥ 20.000	48	23.818	
	20%	35%	0,56
Total	245	67.377	

100% 100%

- **Riscos absolutos, riscos relativos, diferenças de riscos e razões de azares.** A análise de estudos sobre **testes prognósticos** é semelhante àquela para outros estudos de coorte. Se todos os indivíduos em um estudo sobre um teste prognóstico forem seguidos por um determinado período de tempo (p. ex., três anos) com poucas perdas no seguimento, os resultados poderão ser sumarizados com riscos absolutos, riscos relativos e diferenças de riscos. Especialmente quando o segu imento é completo e de curta duração, os resultados de testes prognósticos são às vezes sumarizados da mesma forma que os testes diagnósticos, usando sensibilidade, especificidade, valor preditivo, razões de verossimilhança e curvas ROC. Por outro lado, quando os sujeitos do estudo são seguidos por períodos variáveis de tempo, é preferível usar uma técnica de análise de sobrevida que leve em consideração a duração do seguimento e estime as razões de azares (*hazard ratios*) (8).

- **Melhora líquida após a reclassificação (net reclassification improvement).** Para testes ou biomarcadores novos que buscam predizer a ocorrência futura de eventos de doença, é importante quantificar o quanto os novos testes acrescentam aos modelos de predição existentes. Uma forma de fazer isso é ver o quanto eles aumentam a área sob a curva ROC, mas essas mudanças frequentemente são pequenas, mesmo para preditores bem estabelecidos, sendo difícil inferir a partir desses achados quais mudanças poderiam ocorrer nas decisões clínicas e nos desfechos dos pacientes (9, 10). Uma abordagem mais direta, que é especialmente útil quando os limiares para iniciar tratamento são bem estabelecidos, é examinar com que frequência um modelo ou regra de predição clínica que inclui um novo teste altera a classificação dos pacientes de uma categoria de risco (e decisão terapêutica) para outra, comparado ao modelo antigo. Se o novo teste melhorar a predição, mais sujeitos que desenvolvem o desfecho ("casos") deverão mudar para cima, para uma categoria de maior risco, do que mudar para baixo, para uma categoria de menor risco; o oposto deveria ser verdadeiro para aqueles que não desenvolvem o desfecho ("controles"): o seu risco

deveria mudar para baixo em mais sujeitos do que mudar para cima. A **melhoria líquida após a reclassificação (net reclassification improvement, NRI)** quantifica essas diferenças como mostrado a seguir (11):

$$NRI = P(cima|caso) – P(baixo|caso) + P(baixo|controle) – P(cima|controle)$$

onde P(cima|caso) é a proporção de casos nos quais o modelo com o novo marcador fez com que o sujeito mudasse para uma categoria de maior risco e os demais termos são definidos de forma semelhante. Por exemplo, Shepherd et al. (12) mostraram que acrescentar o volume fibroglandular calculado pela mamografia (isto é, o volume de tecido mamário em risco para malignidade) a um modelo que incluía os fatores de risco clínicos tradicionais melhorou a predição subsequente de câncer de mama ou de carcinoma ductal *in situ*, com um NRI de 21% (P = 0,0001).

ESTUDOS PARA DESENVOLVER REGRAS DE PREDIÇÃO CLÍNICA

Os estudos para desenvolver novas **regras de predição clínica** diferem dos estudos sobre testes (ou regras de predição clínica) já existentes, uma vez que o objetivo é melhorar as decisões clínicas por meio de métodos matemáticos para desenvolver um novo teste composto, em vez de simplesmente avaliar um teste já existente.

Os **sujeitos** nesses estudos devem ser semelhantes àqueles para os quais a regra será aplicada. As regras de predição clínica têm maior utilidade quando ajudam a orientar uma decisão clínica específica, como a decisão de iniciar estatinas (que se baseia no Escore de Risco de Framingham). Dessa forma, os sujeitos devem ser aqueles para os quais será necessário tomar essa decisão clínica, especialmente aqueles para os quais a decisão é atualmente difícil ou incerta (13). Muitos estudos para desenvolver regras de decisão clínica incluem sujeitos de um único centro, mas quando a regra é desenvolvida usando dados de múltiplos centros a

capacidade de generalização costuma ser maior.

Os métodos matemáticos utilizados geralmente envolvem uma técnica multivariada para selecionar variáveis preditoras candidatas e combinar seus valores para gerar uma predição. As variáveis candidatas devem incluir todas as **variáveis preditoras** conhecidas e plausíveis que podem ser medidas de forma fácil, confiável e barata. Um modelo multivariado, como a **regressão logística** ou o **modelo (de azares proporcionais) de Cox**, pode quantificar a contribuição independente de cada variável preditora candidata para a predição do desfecho. Aquelas variáveis mais fortemente ou mais consistentemente associadas com o desfecho podem ser incluídas na regra, e pontos podem ser designados para diferentes valores das variáveis preditoras, conforme os coeficientes que aparecem no modelo. Por exemplo, Wells e colaboradores (14) fizeram uma análise de regressão logística com 40 potenciais preditores clínicos de embolia pulmonar para gerar um escore de predição com base em apenas 7 variáveis (Tabela 12.4). Esse escore agora popular é usado para designar uma probabilidade pré-teste de embolia pulmonar, de modo a orientar futuras decisões sobre realizar ou não testes adicionais e a auxiliar na interpretação de seus resultados (15).

TABELA 12.4 Exemplo de uma regra de predição clínica (para embolia pulmonar) derivada a partir de uma análise de regressão logística (14)

CARACTERÍSTICA CLÍNICA	PONTOS
História prévia de embolia pulmonar ou de trombose venosa profunda	+ 1,5
Frequência cardíaca > 100 batimentos por minuto	+ 1,5
História recente de cirurgia ou imobilização (nos últimos 30 dias)	+ 1,5
Sinais clínicos de trombose venosa profunda	+ 3
Diagnóstico alternativo menos provável do que embolia pulmonar	+ 3
Hemoptise (escarrar sangue)	+ 1
Câncer (tratado nos últimos seis meses)	+ 1
PROBABILIDADE CLÍNICA ESTIMADA DE EMBOLIA PULMONAR (15)	**ESCORE TOTAL**
Baixa (Probabilidade ~1%-2%)	0-1
Intermediária (Probabilidade ~16%)	2-6
Elevada (Probabilidade ~40%)	≥ 7

Uma técnica alternativa, que não requer modelagem e é útil para gerar

regras com elevada sensibilidade, é o **particionamento recursivo**, ou **análise de árvore de classificação e regressão (Classification and Regression Tree, CART)**. Por meio dessa técnica, cria-se uma árvore que faz uma série de perguntas do tipo sim/não, levando o usuário por diferentes caminhos conforme as respostas às perguntas. No final de cada caminho haverá uma probabilidade estimada do desfecho. A árvore pode ser desenvolvida para ter uma elevada sensibilidade, instruindo o pacote de análise estatística a dar uma penalidade maior para falsos-negativos do que para falsos-positivos. Um exemplo de uma árvore dessas, usada para predizer meningite bacteriana em adultos com meningite (16), é mostrado na Figura 12.2.

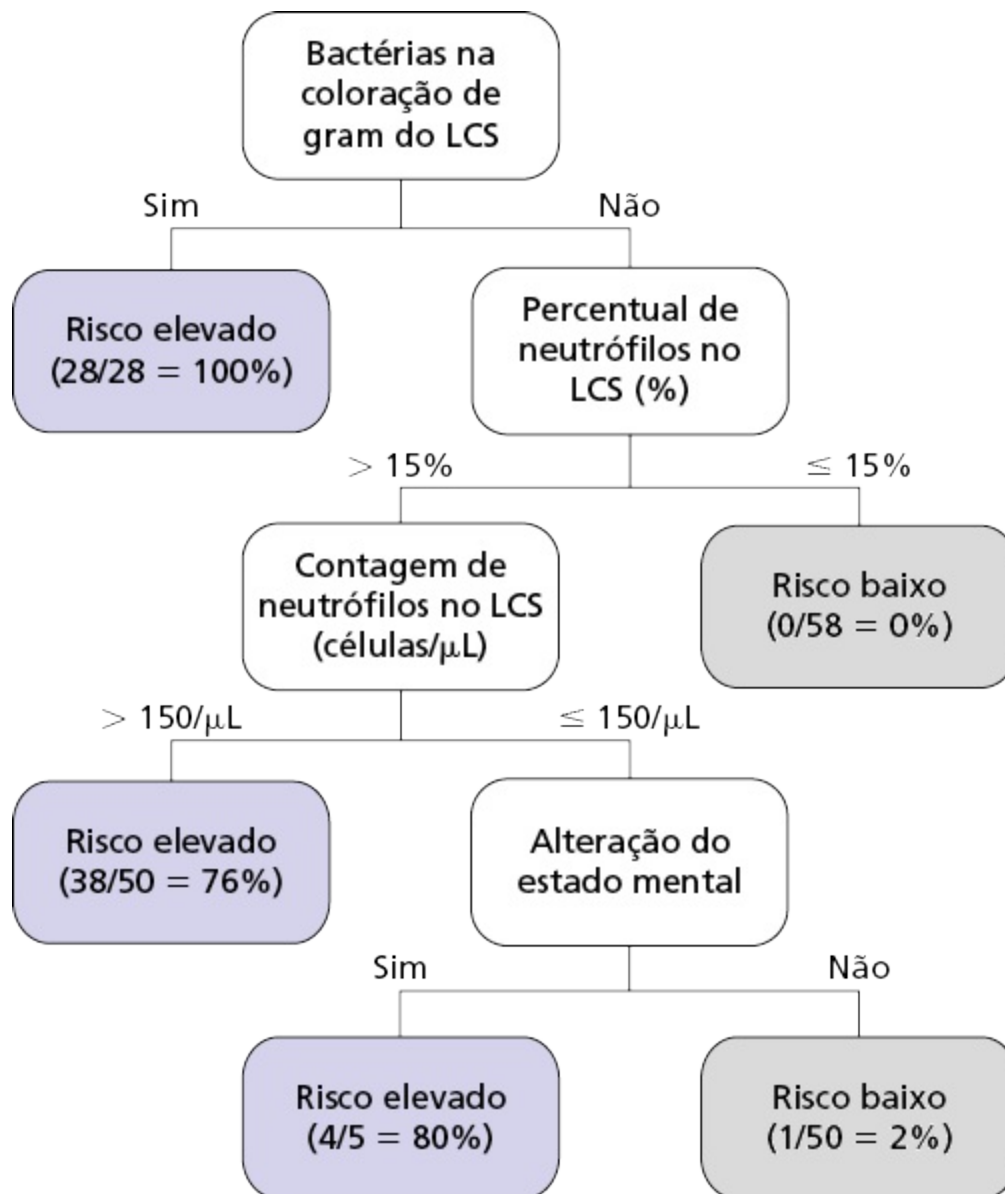



FIGURA 12.2 Exemplo de uma análise de árvore de classificação e regressão (Classification and Regression Tree, CART) para distinguir meningite bacteriana da viral em adultos (16). As caixas brancas servem para dividir os sujeitos em indivíduos com alto risco para meningite bacteriana (caixas azuis) e indivíduos com baixo risco (caixas cinzas); os números mostram as proporções com meningite bacteriana[5] nos "ramos terminais" vermelhos e verdes da árvore. LCS, líquido cerebrospinal.

Independentemente do método escolhido para desenvolver a regra, é importante que ela seja **validada** em um grupo de pacientes diferente daquele no qual ela foi derivada. Um motivo para isso é evitar o **sobreajuste** (isto é, aproveitar-se da tendência em uma única amostra de

o erro aleatório aumentar a força preditiva de alguns fatores). Pode-se lidar com o sobreajuste dividindo a coorte em conjuntos de dados para **derivação** (em média de 50 a 67% da amostra) e **validação**, e usar os dados da coorte de validação para testar a regra derivada a partir da coorte de derivação. Entretanto, essa estratégia valida a regra apenas em uma população muito semelhante àquela de onde ela foi derivada (isto é, lida apenas com a validade interna). Para abordar a validade externa, é importante avaliar se a regra desempenha bem em populações diferentes (“validação prospectiva”) (17).

ESTUDOS SOBRE O EFEITO DOS RESULTADOS DO TESTE NAS DECISÕES CLÍNICAS

Um teste pode até ser acurado, mas, se a doença for muito rara, poderá ser tão raro encontrar um resultado positivo que quase nunca valerá a pena realizá-lo. Outros testes podem não afetar decisões clínicas porque não fornecem informações novas além do que já se sabia (p. ex., a partir da anamnese e do exame físico). Os delineamentos apresentados nesta seção tratam do rendimento de testes diagnósticos e dos seus efeitos nas decisões clínicas.[5]

Tipos de estudos

- **Estudos sobre rendimento diagnóstico.** Os estudos sobre rendimento diagnóstico tratam de questões como as seguintes:
  - Quando um teste é solicitado para uma determinada indicação, com que frequência ele é anormal?
  - É possível predizer resultados anormais a partir de outras informações disponíveis no momento do teste?
  - Em que grupo(s) de pacientes o teste tem o maior ou menor valor?
  - O que acontece com os pacientes com resultados anormais? Os benefícios superam os danos?

Os estudos sobre rendimento diagnóstico estimam a proporção de

testes positivos em pacientes com uma determinada indicação para a realização do teste. Infelizmente, demonstrar que um teste é frequentemente positivo não é suficiente para indicar que ele deve ser realizado. No entanto, um estudo sobre rendimento diagnóstico que demonstra que um teste é quase sempre negativo pode ser suficiente para questionar sua utilidade para essa indicação.

Por exemplo, Siegel e colaboradores (18) estudaram o rendimento da coprocultura em pacientes hospitalizados com diarreia. Embora nem todos os pacientes com diarreia façam coprocultura, parece lógico pressupor que aqueles que fazem têm maior probabilidade de apresentar cultura positiva do que aqueles que não fazem. Apenas 40 (2%) das 1.964 coproculturas foram positivas. Além disso, nenhuma das coproculturas positivas era proveniente dos 997 pacientes internados no hospital há mais de três dias. Como é pouco provável que uma coprocultura negativa afete o manejo desses pacientes com baixa probabilidade de diarreia bacteriana, os autores concluíram que a coprocultura tem pouco valor em pacientes com diarreia internados há mais de três dias.

- **Estudos do tipo antes/depois sobre tomada de decisões clínicas.** Esses delineamentos avaliam o efeito direto que o resultado de um teste tem sobre decisões clínicas e geralmente comparam aquilo que os clínicos fazem (ou dizem que fariam) antes e depois de obterem os resultados de um teste diagnóstico. Por exemplo, Carrico e colaboradores (19) estudaram prospectivamente o valor da ultrassonografia abdominal em 94 crianças com dor aguda no abdome inferior. Os médicos que solicitaram as ultrassonografias foram orientados a registrar sua impressão diagnóstica e a indicar o tratamento como se não houvesse ultrassonografia disponível. Após a ultrassonografia ter sido realizada e o seu resultado informado aos médicos, estes foram questionados novamente sobre suas impressões. A informação fornecida pela ultrassonografia alterou o plano de tratamento inicial em 46% dos pacientes.

É claro que, como discutido mais adiante, a alteração de uma decisão clínica não quer dizer que o paciente será necessariamente beneficiado e, de fato, algumas decisões alteradas podem ser mostrar danosas. Os

estudos que mostram efeitos sobre decisões clínicas têm maior utilidade quando há clareza sobre a história natural da doença e sobre a eficácia do tratamento. No exemplo anterior, provavelmente haveria benefício em alterar a decisão de "alta hospitalar" para "laparoscopia" em crianças com apendicite, ou de "laparoscopia" para "observar" em crianças com dor abdominal não específica.

ESTUDOS SOBRE FACTIBILIDADE, CUSTOS E RISCOS DE TESTES

Uma outra área importante da pesquisa clínica está relacionada aos aspectos práticos do teste diagnóstico. Que proporção dos pacientes irá retornar pelo correio o resultado do teste de Mantoux para tuberculose? Quais são os efeitos médicos de testes de rastreamento falso-positivos em recém-nascidos e quais os efeitos psicológicos nos pais? Que proporção das colonoscopias são complicadas por perfuração colônica?

Delineamentos

Os estudos sobre **factibilidade**, **custos** e **riscos** de testes diagnósticos são geralmente descritivos. Sua estratégia de amostragem é importante porque os testes costumam variar entre indivíduos e instituições que os aplicam, assim como entre os pacientes que os recebem.

Uma opção simples e direta é estudar todos aqueles que recebem o teste, como, por exemplo, em um estudo sobre o índice de retorno dos resultados do teste de Mantoux para tuberculose. Por outro lado, para certas questões, os sujeitos no estudo podem ser apenas aqueles cujos resultados foram positivos ou falso-positivos. Por exemplo, Bodegard e colaboradores (20) estudaram famílias de lactentes que haviam tido resultado falso-positivo em um teste de rastreamento para hipotireoidismo em recém-nascidos e descobriram que as preocupações com a saúde do bebê permaneceram por pelo menos seis meses em quase 20% das famílias.

Não são apenas os resultados falso-positivos que podem causar **efeitos adversos**; o próprio processo de testagem pode causá-los. Por exemplo, Rutter e colaboradores (21) usaram um prontuário médico eletrônico para realizar um estudo de coorte retrospectiva sobre os efeitos adversos

graves (perfuração, hemorragia e diverticulite aguda) nos 30 dias após uma colonoscopia em pacientes do Group Health Cooperative de Puget Sound.

Análise

Os resultados desses estudos podem ser sumarizados com estatísticas simples como médias e desvios-padrão, medianas, faixas de variação e distribuições de frequências. Variáveis dicotômicas, como a ocorrência de efeitos adversos, podem ser sumarizadas com proporções e seus intervalos de confiança (IC) de 95%. Por exemplo, no estudo mencionado, Rutter e colaboradores (21) relataram perfurações em 21/43.456 colonoscopias; isso corresponde a 0,48 por 1.000, com um intervalo de confiança de 0,30 a 0,74 por 1.000.

Em geral, não há uma divisão nítida para distinguir os testes em factíveis e não factíveis ou em testes que tenham ou não tenham risco inaceitavelmente alto de efeitos adversos. Por esse motivo, é útil especificar na etapa de delineamento os critérios para se decidir se o teste é aceitável. Que proporção de seguimento completo deveria ser considerada insuficiente? Que taxa de complicações deveria ser considerada elevada?

ESTUDOS SOBRE O EFEITO DO TESTE NOS DESFECHOS

A melhor forma de determinar o valor de um teste médico é verificar se os pacientes testados têm um desfecho clínico melhor (p. ex, sobrevivem mais ou têm melhor qualidade de vida) do que aqueles que não foram testados. Ensaios clínicos randomizados são o delineamento ideal para isso, porém ensaios clínicos sobre testes diagnósticos costumam ser de difícil execução. Portanto, o valor dos testes médicos é geralmente estimado a partir de estudos observacionais. A principal diferença entre os delineamentos descritos nesta seção e os delineamentos experimentais e observacionais discutidos em outras partes deste livro é que a variável preditora aqui é a *realização de um teste*, e não um tratamento, um fator de risco ou o resultado de um teste.

Delineamentos

É pouco provável que o teste, por si só, traga algum benefício direto para a saúde do paciente. O paciente apenas se beneficia quando o resultado de um teste leva a intervenções preventivas ou terapêuticas eficazes (22). Portanto, uma limitação importante dos estudos sobre desfechos de testes diagnósticos é que a variável preditora em estudo não é apenas um teste (p. ex., teste de sangue oculto nas fezes), mas todo o cuidado médico que o acompanha (p. ex., procedimentos para o seguimento de resultados anormais, colonoscopia, etc.).

O ideal é que a variável de desfecho em estudos como esses seja uma medida de morbidade ou mortalidade, e não apenas um diagnóstico ou estágio da doença. Por exemplo, não basta mostrar que homens rastreados para câncer de próstata têm uma proporção maior de cânceres diagnosticados precocemente para mostrar o valor do rastreamento (23, 24). Muitos desses cânceres não teriam causado nenhum problema se não tivessem sido detectados.

Os desfechos devem ser suficientemente abrangentes para incluir efeitos adversos plausíveis do teste e do tratamento, podendo incluir efeitos psicológicos e médicos do teste. Portanto, um estudo sobre o rastreamento com o antígeno prostático específico para câncer de próstata deveria incluir impotência ou incontinência relacionadas ao tratamento, além da morbidade e mortalidade relacionadas ao câncer. Quando o número de pessoas testadas é muito maior do que aqueles com provável benefício (e isso geralmente é o caso), desfechos adversos menos graves nos indivíduos sem a doença se tornam importantes, uma vez que ocorrerão com maior frequência. Embora, por um lado, encontrar resultados negativos para o teste seja reconfortante ou mesmo um alívio para alguns pacientes (25), por outro lado, os efeitos psicológicos do rótulo ou de resultados falso-positivos, perdas de seguros de saúde e efeitos colaterais problemáticos (mas não fatais) de medicamentos ou procedimentos cirúrgicos preventivos podem sobrepujar os benefícios pouco frequentes (24).

- **Estudos observacionais.** Estudos observacionais são geralmente mais rápidos, mais fáceis e menos dispendiosos do que ensaios clínicos. No entanto, também têm desvantagens consideráveis, especialmente porque os pacientes testados tendem a diferir dos não testados em pontos

importantes que poderiam estar relacionados ao risco de uma doença ou ao seu prognóstico. Por exemplo, indivíduos que recebem o teste poderiam ter um risco relativamente *baixo* de desfecho adverso, porque aqueles que se voluntariam para testes e tratamentos médicos costumam ser mais saudáveis do que a média, um exemplo de **viés de voluntariado**. Por outro lado, os indivíduos testados poderiam também ter um risco relativamente *alto*, pois os pacientes têm maior chance de serem testados quando eles ou seus médicos têm motivos para estarem preocupados com uma doença, um exemplo do **confundimento pela indicação** para o teste (Capítulo 9).

Um outro problema comum em estudos observacionais sobre a realização de testes é a falta de padronização e documentação das intervenções ou das mudanças nos procedimentos posteriores a resultados positivos. Se um teste não melhorou o desfecho em um determinado ambiente, pode ser porque o seguimento dos resultados anormais foi deficiente, porque os pacientes não aderiram à intervenção planejada ou porque a intervenção usada no estudo não foi a ideal.

- **Ensaios clínicos.** O delineamento mais rigoroso para avaliar o benefício de um teste diagnóstico é um ensaio clínico no qual os sujeitos são alocados aleatoriamente para receber ou não o teste. Presumivelmente, o resultado do teste é usado para orientar o manejo clínico do paciente. Medem-se, então, uma série de desfechos, que são comparados nos dois grupos. **Ensaios clínicos randomizados** minimizam ou eliminam o confundimento e o viés de seleção e permitem a aferição de todos os desfechos relevantes, como mortalidade, morbidade, custo e satisfação. A padronização do processo de testar e intervir permite que outros reproduzam os mesmos resultados.

Infelizmente, os ensaios clínicos randomizados sobre testes diagnósticos muitas vezes **não são práticos**, especialmente para testes diagnósticos já usados no cuidado de pacientes doentes. Os ensaios clínicos randomizados são mais factíveis e importantes quando avaliam testes que poderão ser usados em um grande número de indivíduos aparentemente saudáveis, como novos testes de rastreamento.

Esses ensaios clínicos randomizados, no entanto, levantam muitos

questionamentos éticos sobre barrar o uso de testes de potencial valor. Para minimizar esse problema ético, em vez de alocar aleatoriamente os sujeitos para um teste, pode-se designar aleatoriamente que alguns indivíduos recebam uma intervenção que aumente o uso do teste, como lembretes enviados pelo correio e apoio para a marcação das consultas. A análise principal ainda deve seguir a regra da análise por "**intenção de tratar**" – isto é, o grupo inteiro que foi randomizado para receber a intervenção deve ser comparado à totalidade do grupo de comparação. No entanto, a análise por intenção de tratar tende a criar um viés conservador; a eficácia observada da intervenção subestima a eficácia real do teste, uma vez que alguns sujeitos no grupo-controle receberam o teste e alguns sujeitos no grupo de intervenção, não. Esse problema pode ser resolvido em análises secundárias que incluem as taxas de realização dos testes em ambos os grupos e pressupõem que toda a diferença nos desfechos entre os grupos resulte da diferença na frequência de uso dos testes. Os benefícios reais da testagem resultantes da intervenção podem, então, ser estimados algebricamente (8, 28).

EXEMPLO 12.1 Um estudo observacional elegante sobre um teste de rastreamento

Selby e colaboradores (26) realizaram um estudo de caso-controle aninhado no Kaiser Permanente Medical Care Program para avaliar se o rastreamento com sigmoidoscopia reduz o risco de morte por câncer de cólon. Eles compararam as taxas de realização prévia de sigmoidoscopia em pacientes que morreram por câncer de cólon com controles que não morreram por essa doença. Encontraram uma razão de chances ajustada de 0,41 (IC 95% 0,25 a 0,69), sugerindo que a sigmoidoscopia resultou em uma redução de quase 60% na mortalidade por câncer de reto e cólon distal.

Um potencial problema é que pacientes que fazem sigmoidoscopia podem diferir em aspectos importantes daqueles que não fazem, e essas diferenças podem estar associadas à diferença encontrada na mortalidade esperada para o câncer de cólon. Para abordar esse

possível confundimento, Selby e colaboradores examinaram a aparente eficácia da sigmoidoscopia na prevenção da mortalidade por neoplasias do cólon proximal, que estão além do alcance do sigmoidoscópio. Se os pacientes que fizeram sigmoidoscopia tivessem menor probabilidade de morrer de câncer de cólon por outros motivos, a sigmoidoscopia iria protegê-los também dessas neoplasias. No entanto, a sigmoidoscopia não teve efeito sobre a mortalidade por câncer de cólon proximal (razão de chances ajustadα = 0,96, IC de 95%, 0,61 a 1,50), sugerindo que o confundimento não era a causa da aparente redução na mortalidade por câncer do cólon distal. Especificar desfechos alternativos (*a priori*!) que se imagina não estarem associados ao preditor de interesse (câncer do cólon proximal, nesse caso), e então mostrar que eles realmente não estão associados, pode ajudar muito a fortalecer a inferência causal (27).

Análise

A análise de estudos que abordam o efeito da realização do teste sobre o desfecho deve ser adequada ao delineamento usado – razões de chances para estudos de caso-controle e razões de riscos ou razões de azares (*hazard ratios*) para estudos de coorte ou ensaios clínicos. Uma forma conveniente de relatar os dados é projetar os resultados do procedimento usado no teste para uma coorte grande (p. ex., 100 mil) e listar o número de testes iniciais, testes de seguimento, indivíduos tratados, efeitos colaterais do tratamento, custos e mortes nos grupos que realizaram e não realizaram o teste.

DIFICULDADES NO DELINEAMENTO OU NA ANÁLISE DE ESTUDOS SOBRE TESTES DIAGNÓSTICOS

Como nos outros tipos de pesquisa clínica, as concessões que são feitas no delineamento de estudos sobre testes diagnósticos podem ameaçar a validade dos resultados, e erros na análise podem comprometer sua interpretação. A seguir são apresentados alguns dos erros mais comuns e graves, juntamente com os passos que deveriam ser seguidos para evitá-

los.

Tamanho de amostra inadequado

Se o desfecho em um estudo sobre um teste diagnóstico for comum, provavelmente será factível obter um tamanho de amostra adequado. Quando a doença ou desfecho testado são raros, um número muito grande de pessoas pode ser necessário. Muitos testes laboratoriais, por exemplo, não são caros, e um rendimento de 1% ou menos poderia justificar sua adoção, especialmente se capazes de diagnosticar uma doença grave e tratável.

Por exemplo, Sheline e Kehr (29) revisaram retrospectivamente testes laboratoriais de rotina para admissão, incluindo o teste do Venereal Disease Research Laboratory (VDRL), usado para identificar sífilis, em 252 pacientes psiquiátricos, e observaram que os testes haviam identificado um único paciente com sífilis anteriormente insuspeitada. Se os sintomas psiquiátricos do paciente realmente foram causados pela sífilis, é difícil afirmar que não valeu a pena gastar os 3.186 dólares em VDRLs para fazer esse diagnóstico. Mas se a prevalência real de sífilis insuspeitada fosse próxima à de 0,4% encontrada nesse estudo, um estudo desse tamanho de amostra poderia facilmente não ter encontrado nenhum caso.

Exclusão inadequada

Quando se calculam proporções, não é adequado excluir pacientes do numerador sem excluí-los do denominador. Por exemplo, em um estudo sobre testes laboratoriais de rotina em pacientes atendidos em serviço de emergência com um primeiro episódio de crise convulsiva (30), 11 dos 136 pacientes (8%) tinham uma anormalidade laboratorial passível de correção (p. ex., hipoglicemia) como causa desse episódio convulsivo. Em 9 dos 11 pacientes, no entanto, suspeitou-se da anormalidade a partir da história clínica ou do exame físico. Os autores, portanto, relataram que apenas 2 dos 136 pacientes (1,5%) tinham anormalidades não suspeitadas pela história clínica ou pelo exame físico. No entanto, se todos os pacientes com anormalidades suspeitadas foram excluídos do numerador, pacientes semelhantes deveriam ter sido excluídos também do denominador. Dessa forma, o número correto do denominador para essa

proporção não é todos os 136 pacientes testados, mas somente aqueles em que não se suspeitou da presença de anormalidades laboratoriais a partir da história clínica ou do exame físico.

Excluindo resultados limítrofes ou não interpretáveis

Em algumas situações, um teste pode não fornecer qualquer resposta, como, por exemplo, quando o ensaio laboratorial não funcionou, a amostra testada deteriorou-se ou o resultado do teste ficou na penumbra, não sendo nem positivo nem negativo. Em geral, não se devem ignorar esses problemas, mas a estratégia para lidar com eles depende da questão de pesquisa e do delineamento usado. Em estudos sobre custos ou inconveniências dos testes, falhas no teste constituem resultados claramente importantes.

É importante contabilizar os pacientes com estudos de imagem "não conclusivos" ou com um resultado limítrofe em um teste. Isso pode alterar um teste originalmente dicotômico (positivo, negativo) para um teste ordinal (positivo, negativo e indeterminado). Curvas ROC podem, então, ser plotadas, e razões de verossimilhança, calculadas para resultados "indeterminados", positivos e negativos.

Viés de verificação: aplicação de um único padrão-ouro a um grupo selecionado de pacientes

Uma estratégia comum de amostragem para estudos sobre testes médicos é estudar (prospectiva ou retrospectivamente) pacientes que são testados para uma doença e que também são submetidos ao padrão-ouro para estabelecer o diagnóstico. No entanto, isso pode introduzir um problema se os achados que estão sendo estudados também estiverem sendo usados para decidir quem recebe o padrão-ouro. Por exemplo, considere um estudo sobre os preditores de fraturas em crianças que vão ao serviço de emergência com lesões no tornozelo, no qual foram incluídas apenas crianças que fizeram raio X para essas lesões. Caso aquelas crianças com um determinado achado (p. ex., edema no tornozelo) tivessem maior probabilidade de realizar um exame de raios X, isso poderia afetar a sensibilidade e a especificidade do edema de tornozelo como teste para fratura. Esse viés, denominado **viés de verificação**, é ilustrado com um exemplo numérico no Apêndice 12B. O viés de verificação pode ser

evitado usando-se critérios rigorosos para a aplicação do padrão-ouro que não incluem o teste ou achado que está sendo estudado. Se essa estratégia não for aplicável, é possível estimar e corrigir para o viés de verificação se o padrão-ouro puder ser aplicado a uma amostra aleatória das pessoas que tiverem resultado negativo no teste.

Viés de verificação diferencial: uso de padrões-ouro diferentes para aqueles com resultados positivos e negativos

Outra estratégia é usar um padrão-ouro diferente naquelas pessoas para as quais o padrão-ouro frequente não está indicado. Por exemplo, os pacientes com lesão no tornozelo que não fizeram radiografia poderiam ser contatados por telefone poucas semanas após a lesão e classificados como não tendo sofrido fratura se eles se recuperaram bem. Entretanto, isso pode levar a um **viés de verificação diferencial**, também chamado de **viés de duplo padrão-ouro** (31). Esse viés pode ocorrer sempre que o padrão-ouro diferir entre as pessoas com resultados positivos e negativos. No estudo mencionado sobre mamografia (5), o padrão-ouro para mulheres com resultados positivos era uma biópsia, o padrão-ouro para aquelas com resultados negativos era um período de seguimento para avaliar se elas iriam apresentar manifestações evidentes de câncer de mama ao longo do ano seguinte. Ter dois padrões-ouro para uma doença constitui um problema, se os mesmos não concordarem sempre um com o outro, como ocorreria se o câncer detectado pela biópsia após uma mamografia positiva não se tornasse evidente após um ano de seguimento em pessoas com mamografia negativa.

Um outro exemplo é um estudo sobre o uso da ultrassonografia para diagnosticar intussuscepção em crianças (32). Todas as crianças com resultado positivo para intussuscepção na ultrassonografia foram submetidas ao padrão-ouro, um enema contrastado. Por outro lado, a maioria das crianças com resultado negativo na ultrassonografia foram observadas na emergência, sendo a intussuscepção descartada com base nos achados clínicos. Para casos de intussuscepção com resolução espontânea, os dois padrões-ouro iriam fornecer resultados diferentes: o enema contrastado seria positivo, ao passo que o seguimento clínico seria negativo. Um exemplo numérico desse tipo de viés de verificação de duplo padrão-ouro é apresentado no Apêndice 12C.

O viés de verificação diferencial pode ser evitado aplicando o mesmo padrão-ouro a todos os sujeitos. Quando isso não for factível (como no caso do estudo sobre mamografia), os investigadores devem fazer todo o possível para identificar outros estudos que permitam avaliar o quanto esse viés poderia ameaçar a validade do estudo (p. ex., estudos de necropsias para avaliar a prevalência de neoplasias assintomáticas em pacientes que morreram por outras causas em um estudo sobre rastreamento para câncer).

RESUMO

1. A utilidade de **testes médicos** pode ser avaliada usando-se delineamentos que abordem uma série de questões de complexidade crescente (Tabela 12.1). Em geral, **delineamentos observacionais** comuns fornecem **estatísticas descritivas** sobre as características dos testes juntamente com intervalos de confiança.
2. Devem-se selecionar os **sujeitos** para um estudo sobre um teste diagnóstico a partir de pacientes com um **espectro** de doença e de não doença apropriado para a questão de pesquisa, que na maioria dos casos reflete o uso antecipado do teste na prática clínica.
3. Se possível, o investigador deve **cegar** aqueles que irão fazer a interpretação dos resultados e a determinação do padrão-ouro, evitando que tenham acesso a outras informações sobre o paciente testado.
4. Em geral, medir a **reprodutibilidade**, incluindo a variabilidade **inter** e **intraobservador**, é um bom passo inicial para se avaliar um teste.
5. Em estudos sobre a **acurácia de testes**, é necessário escolher um **padrão-ouro** para que se determine se um paciente tem ou não a doença ou desfecho em estudo.
6. Os resultados de estudos sobre a acurácia de testes diagnósticos podem ser sumarizados usando-se **sensibilidade**, **especificidade**, **valores preditivos**, **curvas ROC** e **razões de verossimilhança**. Estudos sobre o valor de testes prognósticos podem ser sumarizados usando-se **razões de risco**, **razões de azares** ou **melhora líquida após a reclassificação.**
7. Os estudos para desenvolver **novas regras de predição clínica** estão

sujeitos a problemas de **sobreajuste** e baixa **reprodutibilidade**, tornando necessário que as novas regras sejam **validadas** em outras amostras populacionais.

8. O delineamento mais rigoroso para estudar a utilidade de um teste diagnóstico é o **ensaio clínico**, onde os sujeitos são randomizados para receberem ou não o teste. Os desfechos de interesse incluem a **mortalidade**, a **morbidade**, os **custos** e a **qualidade de vida**.
9. Se os ensaios clínicos não forem éticos ou factíveis, pode ser útil realizar **estudos observacionais sobre benefícios, danos e custos**, com atenção adequada a possíveis **vieses** e **variáveis confundidoras**.

APÊNDICE 12A
Cálculo de kapa para medir a concordância interobservador

Considere dois observadores que buscam identificar um ritmo de galope por B4 na ausculta cardíaca (Tabela 12A.1). A observação é registrada como presente ou ausente. A medida mais simples de concordância interobservador é a proporção de observações em que os dois observadores concordam. Essa proporção pode ser obtida somando-se os números da diagonal entre o quadrante superior esquerdo e o quadrante inferior direito e dividindo-se esse resultado pelo total de observações realizadas. Nesse exemplo, entre 100 pacientes observados, havia 10 em que ambos os observadores ouviram um galope, e 75 em que nenhum ouviu, resultando em (10 + 75) / 100 = 85%.

TABELA 12.A.1 Concordância interobservador sobre a presença de um ritmo de galope por B4

	OBSERVADOR 1 OUVIU RITMO DE GALOPE	OBSERVADOR 1 NÃO OUVIU RITMO DE GALOPE	TOTAL, OBSERVADOR 2
Observador 2 ouviu ritmo de galope	10	5	15
Observador 2 não ouviu ritmo de galope	10	75	85
Total, observador 1	20	80	100

Quando as observações não são distribuídas uniformemente entre as categorias (p. ex., quando a proporção de "anormais" em um teste dicotômico é muito diferente de 50%), ou quando há mais de duas categorias, outra medida de concordância interobservador, denominada kapa (κ), é às vezes utilizada. Kapa mede o grau de concordância além do que seria esperado pelo mero acaso, dados os "*valores marginais observados*" (i. e., os totais para as linhas e colunas). Os valores de kapa variam de –1 (discordância completa) a 1 (concordância completa). Um kapa de 0 indica que o grau de concordância é exatamente igual àquele esperado pelos totais das linhas e colunas. O κ é estimado da seguinte forma:

$$\kappa = \frac{\text{Concordância observada (\%) – concordância esperada (\%)}}{\text{100\% – Concordância esperada (\%)}}$$

A proporção “esperada” para cada célula é simplesmente a proporção observada naquela linha (i. e., o total da linha dividido pelo tamanho da amostra) multiplicada pela proporção observada naquela coluna (i. e., o total da coluna dividido pelo tamanho da amostra). A concordância esperada é obtida somando-se as proporções esperadas nas células na diagonal correspondente à concordância dos observadores.

Por exemplo, na Tabela 12A.1, os observadores apresentaram alta concordância: eles concordaram 85% das vezes. Mas qual foi a concordância observada além daquela esperada pelos valores totais marginais? Apenas pelo acaso (dados os valores marginais observados), eles teriam concordado em 71% das vezes: (20% × 15%) + (80% × 85%) = 71%. Como a concordância observada era de 85%, o kapa é (85% – 71%)/(100% – 71%) = 0,48, uma concordância apreciável, mesmo que expressa de forma menos impressionante do que 85%.

Quando há mais de duas categorias de resultados de testes, é importante distinguir entre variáveis ordinais, que são intrinsecamente ordenadas, e variáveis nominais, que não são. Para variáveis ordinais, o kapa, como calculado acima, não capta toda a informação dos dados, porque não dá crédito parcial para resultados mais próximos, mesmo que discordantes. Para dar crédito para concordâncias parciais, deve-se usar um kapa ponderado. (Ver Newman e Kohn [31] para uma discussão mais detalhada.)

APÊNDICE 12B
Exemplo numérico de viés de verificação

Considere dois estudos sobre o edema no tornozelo como preditor de fraturas em crianças com lesões no tornozelo. O primeiro estudo utilizou uma **amostra consecutiva** de 200 crianças. Nesse estudo, todas as crianças com lesões no tornozelo foram submetidas a raios X, independentemente do edema. A sensibilidade e a especificidade do edema no tornozelo são de 80 e 75%, respectivamente, conforme mostrado na Tabela 12B.1:

TABELA 12B.1 Edema no tornozelo como preditor de fratura utilizando uma amostra consecutiva

	COM FRATURA	SEM FRATURA
Com edema	32	40
Sem edema	8	120
Total	40	160
	Sensibilidade = 32/40 = 80%	Especificidade = 120/160 = 75%

O segundo estudo usou uma amostra **selecionada**, na qual apenas metade das crianças sem edema no tornozelo foram submetidas a raios X. Portanto, os números na coluna "Sem edema" são reduzidos pela metade. Isso aumenta a sensibilidade aparente de 32/40 (80%) para 32/36 (89%) e diminui a especificidade aparente de 120/160 (75%) para 60/100 (60%), como demonstrado na Tabela 12B.2:

TABELA 12B.2 Viés de verificação: edema no tornozelo como preditor de fraturas usando uma amostra selecionada

	COM FRATURA	SEM FRATURA
Com edema	32	40
Sem edema	4	60
Total	36	100
	Sensibilidade = 32/36 = 89%	Especificidade = 60/100 = 60%

APÊNDICE 12C
Exemplo numérico de viés de verificação diferencial

Os resultados do estudo de Eshed e colaboradores sobre ultrassonografia para diagnosticar intussuscepção (32) são apresentados na Tabela 12C.1:

TABELA 12C.1 Resultados de um estudo sobre ultrassonografia para diagnosticar intussuscepção

	COM INTUSSUSCEPÇÃO	SEM INTUSSUSCEPÇÃO
Ultrassonografia +	37	7
Ultrassonografia –	3	104
Total	40	111
	Sensibilidade = 37/40 = 93%	Especificidade = 104/111 = 94%

Os 104 indivíduos com resultados negativos na ultrassonografia listados como "Sem intussuscepção" na verdade incluíam 86 indivíduos que foram seguidos clinicamente e que não foram submetidos a enema contrastado. Se 10% desses indivíduos (i. e., nove crianças) tivessem intussuscepção com resolução espontânea, mas que ainda assim teria sido identificada por meio de um enema contrastado, e se todos os indivíduos tivessem recebido enema contrastado, essas nove crianças teriam seu resultado alterado de verdadeiro-negativo para falso-negativo, conforme mostrado na Tabela 12C.2:

TABELA 12C.2 Efeito sobre a sensibilidade e a especificidade de nove crianças com intussuscepção com resolução espontânea tivessem recebido o padrão-ouro enema contrastado, em vez de seguimento clínico

	COM INTUSSUSCEPÇÃO	SEM INTUSSUSCEPÇÃO
Ultrassonografia +	37	7
Ultrassonografia –	3 + 9 = 12	104 – 9 = 95
Total	49	102
	Sensibilidade = 37/49 = 76%	Especificidade = 95/102 = 93%

Um efeito semelhante, porém menos pronunciado, teria ocorrido se algumas crianças com resultados positivos no exame tivessem tido intussuscepção com resolução espontânea caso fosse dada a chance (31).

REFERÊNCIAS

1. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307–310.
2. Newman TB, Kohn M. *Evidence-based diagnosis*. New York: Cambridge University Press, 2009:10–38.
3. Tokuda Y, Miyasato H, Stein GH, et al. The degree of chills for risk of bacteremia in acute febrile illness. *Am J Med* 2005;118(12):1417.
4. Sawaya GF, Washington AE. Cervical cancer screening: which techniques should be used and why? *Clin Obstet Gynecol* 1999;42(4):922–938.
5. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of - mammographic accuracy. *J Natl Cancer Inst* 2005;97(5):358–367.
6. Rocker G, Cook D, Sjokvist P, et al. Clinician predictions of intensive care unit mortality. *Crit Care Med* 2004;32(5):1149–1154.
7. Newman TB, Puopolo KM, Wi S, et al. Interpreting complete blood counts soon after birth in newborns at risk for sepsis. *Pediatrics* 2010;126(5):903–909.
8. Vittinghoff E, Glidden D, Shiboski S, et al. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*, 2nd ed. New York: Springer, 2012.
9. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med* 2009;150(11):795–802.
10. Cook NR. Assessing the incremental role of novel and emerging risk factors. *Curr Cardiovasc Risk Rep* 2010;4(2):112–119.
11. Pencina MJ, D’Agostino RB, Sr., D’Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27(2):157–172; discussion 207–212.
12. Shepherd JA, Kerlikowske K, Ma L, et al. Volume of mammographic density and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2011;20(7):1473–1482.
13. Grady D, Berkowitz SA. Why is a good clinical prediction rule so hard to find? *Arch Intern Med* 2011;171(19):1701–1702.
14. Wells PS, Anderson DR, Rodger M, et al. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the SimpliRED D-dimer.*Thromb Haemost* 2000;83(3):416–420.
15. Wells PS, Anderson DR, Rodger M, et al. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. *Ann Intern Med* 2001;135(2):98–107.
16. Tokuda Y, Koizumi M, Stein GH, et al. Identifying low-risk patients for bacterial meningitis in adult patients with acute meningitis. *Intern Med* 2009;48(7):537–543.
17. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277(6):488–494.

18. Siegel DL, Edelstein PH, Nachamkin I. Inappropriate testing for diarrheal diseases in the hospital. *JAMA* 1990;263(7):979–982.
19. Carrico CW, Fenton LZ, Taylor GA, et al. Impact of sonography on the diagnosis and treatment of acute lower abdominal pain in children and young adults. *American Journal of Roentgenology* 1999;172(2):513–516.
20. Bodegard G, Fyro K, Larsson A. Psychological reactions in 102 families with a newborn who has a falsely positive screening test for congenital hypothyroidism. *Acta Paediatr Scand Suppl* 1983;304:1–21.
21. Rutter CM, Johnson E, Miglioretti DL, et al. Adverse events after screening and follow-up colonoscopy. *Cancer Causes Control* 2012;23(2):289–296.
22. Etzioni DA, Yano EM, Rubenstein LV, et al. Measuring the quality of colorectal cancer screening: the importance of follow-up. *Dis Colon Rectum* 2006;49(7):1002–1010.
23. Welch HG. *Should I be tested for cancer? Maybe not, and here's why*. Berkeley, CA: University of California Press, 2004.
24. Welch HG, Schwartz LM, Woloshin S. *Overdiagnosed: making people sick in pursuit of health*. Boston, MA: Beacon Press, 2011.
25. Detsky AS. A piece of my mind. Underestimating the value of reassurance. *JAMA* 2012;307(10): 1035–1036.
26. Selby JV, Friedman GD, Quesenberry CJ, et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer [see comments]. *N Engl J Med* 1992;326(10):653–657.
27. Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? *JAMA* 2013;309(3):241–242.
28. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995;57(1):6–15.
29. Sheline Y, Kehr C. Cost and utility of routine admission laboratory testing for psychiatric inpatients. *Gen Hosp Psychiatry* 1990;12(5):329–334.
30. Turnbull TL, Vanden Hoek TL, Howes DS, et al. Utility of laboratory studies in the emergency department patient with a new-onset seizure. *Ann Emerg Med* 1990;19(4):373–377.
31. Newman TB, Kohn MA. *Evidence-based diagnosis*. New York: Cambridge University Press, 2009: 101–102.
32. Eshed I, Gorenstein A, Serour F, et al. Intussusception in children: can we rely on screening sonography performed by junior residents? *Pediatr Radiol* 2004;34(2):134–137.

---

[1] Uma alternativa é comparar a acurácia da anamnese e do exame físico isoladamente com a acurácia da anamnese e do exame físico somados ao ultrassom.

[2] O kapa pode ser descrito como o grau de concordância além do que seria esperado pelo acaso, mas a estimativa da concordância esperada pelo acaso baseia-se na prevalência da anormalidade determinada por cada observador.

[3] Embora o coeficiente de variação seja frequentemente utilizado, é melhor evitá-lo em estudos sobre a confiabilidade de exames laboratoriais, uma vez que é fortemente influenciado por valores extremos e não permite aos leitores determinar com que frequência as diferenças entre as duas aferições são clinicamente importantes. Os ICs para a diferença média também devem ser evitados, uma vez que sua dependência sobre o tamanho de amostra os torna potencialmente enganadores. Um IC estreito para a diferença média entre as duas aferições não significa que elas geralmente concordem – apenas que a diferença média entre elas está sendo medida de forma precisa. Veja Bland e Altman (1) ou Newman e Kohn (2) para discussão adicional sobre esses assuntos.

[4] Para testes dicotômicos, a razão de verossimilhança para um teste positivo é

$$\frac{\text{Sensibilidade}}{1 - \text{Especificidade}}$$

e a razão de verossimilhança para um teste negativo é

$$\frac{1 - \text{Sensibilidade}}{\text{Especificidade}}$$

Discussões detalhadas sobre como usar as razões de verossimilhança e informações prévias (a probabilidade prévia de doença) para estimar a probabilidade de um paciente ter uma doença após saber o resultado do teste (probabilidade posterior) estão disponíveis no livro de Newman e Kohn (2). A fórmula é

Chances prévias × Razão de verossimilhança = Chances posteriores

onde as chances prévias e posteriores estão relacionadas a suas respectivas probabilidades pela fórmula

$$\text{Chances} = \frac{P}{1 - P}$$

[5] Os dados na figura incluem tanto o conjunto de dados para derivação quanto o conjunto para validação.

CAPÍTULO 13

Pesquisas com dados existentes

Deborah G. Grady, Steven R. Cummings e Stephen B. Hulley

Muitas questões de pesquisa podem ser respondidas de forma rápida e eficiente a partir de dados ou amostras biológicas já coletados. Há três abordagens gerais para utilizar esses recursos existentes. A **análise de dados secundários** é o uso de dados existentes para investigar questões de pesquisa diferentes daquelas para as quais os dados foram originalmente coletados. **Estudos suplementares** acrescentam uma ou mais medidas a um estudo, muitas vezes em um subconjunto de participantes, para responder a uma outra questão de pesquisa. **Revisões sistemáticas** combinam os resultados de vários estudos anteriores sobre uma determinada questão de pesquisa, muitas vezes calculando uma estimativa-sumário de efeito que tem uma precisão maior do que a das estimativas dos estudos individuais. O uso criativo de dados e amostras biológicas existentes é uma forma eficaz para jovens investigadores com recursos limitados começarem a responder a questões de pesquisa importantes, ganharem experiência valiosa em uma área de pesquisa e às vezes encontrarem um achado publicável em um curto espaço de tempo.

VANTAGENS E DESVANTAGENS

As principais **vantagens** de estudos com dados existentes são a economia de tempo e de recursos financeiros. Dessa forma, uma questão de pesquisa que poderia demandar muito tempo e recursos para ser investigada pode ser respondida de **forma rápida** e **barata**. Por exemplo, no banco de dados do Study of Osteoporotic Fractures, em um estudo de coorte

prospectivo originalmente delineado para estudar fatores de risco para fraturas, Yaffe e colaboradores utilizaram medidas repetidas que haviam sido feitas dos níveis de atividade física e da função cognitiva e descobriram que mulheres que caminhavam mais tinham um risco de declínio cognitivo 36% menor do que mulheres que caminhavam menos. (1).

No entanto, os estudos que utilizam dados ou amostras biológicas existentes também têm suas **desvantagens**. A seleção da população a ser estudada e dos dados a serem coletados, a qualidade dos dados coletados e a forma como as variáveis são aferidas e registradas são todas predeterminadas. Os dados existentes podem ter sido coletados de uma população que não é a ideal (p. ex., *apenas homens, em vez de homens e mulheres*), a estratégia de aferição pode não ser a que o investigador gostaria (*história de hipertensão como variável dicotômica, em vez da própria aferição da pressão arterial em mmHg*), e a qualidade dos dados pode ser deficiente (*muitos dados faltantes ou incorretos*). Confundidores e desfechos importantes podem não ter sido medidos ou registrados. Todos esses fatores contribuem para a principal desvantagem de se usar dados existentes: o investigador tem pouco ou **nenhum controle** sobre quais dados foram coletados e sobre como eles foram coletados.

ANÁLISE DE DADOS SECUNDÁRIOS

Dados secundários podem vir de registros de prontuário, arquivos de cobrança, certidões de óbito, bases de dados públicas e muitas outras fontes, porém **estudos prévios** conduzidos na própria instituição do investigador ou em outros locais constituem uma das fontes mais ricas de dados secundários. Muitos estudos coletam mais dados do que o investigador pode analisar, e esses dados podem ser utilizados para documentar resultados interessantes que passaram despercebidos. O acesso a esses dados é geralmente controlado pelo **investigador principal** do estudo ou por um **comitê diretivo**; um investigador iniciante deve, portanto, procurar informações sobre estudos realizados por outros pesquisadores que podem ter feito aferições relevantes para a questão de pesquisa. Uma das melhores formas de um bom mentor ajudar um investigador iniciante é fornecendo informações e acesso a bancos de

dados relevantes. A maioria dos estudos **financiados pelo NIH** deve tornar seus dados disponíveis ao público.

Outras fontes úteis de dados secundários são **bases de dados nacionais** disponíveis para o grande público e que não têm um investigador principal. Bases de dados computadorizadas desse tipo são tão variadas quanto são as razões que levam investigadores a coletar informações. A seguir, apresentaremos diversos exemplos que merecem atenção especial, e os leitores podem localizar outros exemplos nas suas áreas de interesse.

- **Registros de câncer**[1] são mantidos por entidades financiadas pelo governo que coletam estatísticas completas sobre incidência, tratamento e desfechos de câncer em áreas geográficas definidas. Esses registros, atualmente, incluem cerca de um quarto da população norte-americana, mas espera-se que a cobertura geográfica seja ampliada nos próximos anos. Um objetivo desses registros é fornecer dados para investigadores de fora do programa. Os dados de todos os registros estão disponíveis no Surveillance, Epidemiology and End Results (SEER) Program. Por exemplo, investigadores usaram o registro do SEER de diagnósticos de câncer de mama para mostrar que a incidência anual de câncer de mama positivo para receptor de estrogênio diminuiu 13% em mulheres pós-menopáusicas entre 2001 e 2003; essa tendência acompanhou a redução do uso da terapia de reposição hormonal por mulheres pós-menopáusicas, o que sugere que parar de fazer terapia de reposição hormonal diminui o risco de câncer de mama (2).
- **Registros de declarações de óbito** podem ser usados para acompanhar a mortalidade de qualquer coorte. O **National Death Index** inclui todos os óbitos nos Estados Unidos desde 1978. Ele pode ser usado para averiguar o estado vital de sujeitos de um estudo anterior ou de sujeitos que são parte de outra base de dados que inclui variáveis preditoras importantes. Um exemplo clássico é o seguimento de homens com doença coronariana alocados aleatoriamente para ácido nicotínico em alta dose ou placebo com o intuito de baixar o colesterol sérico no Coronary Drug Project. Nenhum estudo havia mostrado benefício de tratar o colesterol elevado sobre mortalidade, e não houve diferença na mortalidade ao final dos cinco anos de tratamento randomizado,

entretanto, um seguimento de nove anos para mortalidade usando o National Death Index revelou um benefício significativo (3). O estado vital de um indivíduo é informação pública, portanto, foi possível completar o seguimento até mesmo para homens que haviam abandonado o estudo.

O National Death Index pode ser usado quando se conhece o nome, a data de nascimento ou o número de seguridade social. Com esse sistema, a averiguação do óbito fica 99% completa, e é possível obter informações adicionais das declarações de óbito (especialmente a causa da morte) a partir de registros estaduais. Em nível estadual e local, muitas jurisdições já dispõem de sistemas computadorizados de estatísticas vitais nos quais dados individuais (como informações das declarações de nascimento e óbito) são lançados conforme vão sendo recebidos.

- **NHANES** (National Health and Nutrition Examination Survey – Inquérito Nacional sobre Saúde e Nutrição)[2] consiste em uma série de inquéritos que avaliam o estado de saúde e nutricional de adultos e crianças nos Estados Unidos. Esses inquéritos empregam seleção aleatória por conglomerados, de base populacional, para identificar uma amostra representativa daquele país. Eles incluem dados autorrelatados (p. ex., demográficos, socioeconômicos, alimentares e de comportamentos relacionados à saúde), exame físico, exames laboratoriais, entre outros. O NHANES fornece estimativas de base populacional sobre a prevalência de doenças, fatores de risco e outras variáveis. Por exemplo, a densidade mineral óssea (DMO) do quadril foi medida em dois desses inquéritos: 1988-1994 e 2005-2006. Os resultados fornecem os valores normais para mulheres e homens de várias raças nos Estados Unidos que são usados para definir "osteoporose" como 2,5 desvios-padrão abaixo da média de DMO para adultos jovens no NHANES (4). Os investigadores também usaram medidas repetidas para mostrar que a DMO está melhorando e a prevalência de osteoporose está diminuindo (5).

Dados secundários podem ser especialmente úteis em estudos que avaliam padrões de utilização de tratamentos médicos e seus desfechos clínicos. Essa abordagem pode complementar as informações fornecidas

por ensaios clínicos randomizados, além de permitir investigar questões de pesquisa que os ensaios clínicos não conseguem responder. Esses tipos de dados existentes incluem **bases eletrônicas de dados administrativos e clínicos**, como as desenvolvidas pelo Medicare, pelo Department of Veterans Affairs, pelos Kaiser Permanente Medical Groups, pelo Duke Cardiovascular Disease Databank, e por **registros** como o San Francisco Mammography Registry e o National Registry of Myocardial Infarction. Dados dessas fontes (muitos dos quais podem ser obtidos na internet) podem ser muito úteis para estudar eventos adversos raros e para avaliar a utilização e a efetividade reais de uma intervenção já demonstrada como eficaz em um ensaio clínico. Por exemplo, o National Registry of Myocardial Infarction foi usado para avaliar fatores de risco para hemorragia intracraniana após o tratamento com um ativador tecidual do plasminogênio recombinante (tPA) em pacientes com infarto do miocárdio. O registro incluía 71.073 pacientes que receberam tPA; 673 tiveram hemorragia intracraniana confirmada por TC ou RM. Uma análise multivariada mostrou que uma dose de tPA superior a 1,5 mg/kg estava significativamente associada ao desenvolvimento de hemorragia intracraniana quando comparada com doses menores (6). Como o risco global de se desenvolver hemorragia intracraniana era de menos de 1%, um ensaio clínico que coletasse dados primários para examinar esse desfecho teria tamanho e custo proibitivos.

Outra contribuição valiosa desse tipo de análise de dados secundários é uma melhor compreensão da diferença entre eficácia e efetividade. O ensaio clínico randomizado é o padrão-ouro para se determinar a **eficácia** de uma terapia em uma população selecionada, sob circunstâncias controladas em contextos clínicos restritos. No “mundo real”, no entanto, os pacientes que são tratados, as escolhas de medicamentos e doses pelo médico e a adesão aos medicamentos pelo paciente costumam ser mais variáveis. Esses fatores podem tornar a aplicação do tratamento na população geral menos efetiva do que o observado em ensaios clínicos. A **efetividade** de tratamentos na prática clínica pode ser avaliada por meio de estudos com dados secundários. Por exemplo, demonstrou-se que a angioplastia primária é superior à terapia trombolítica em ensaios clínicos com pacientes com infarto do miocárdio (7). Contudo, é possível que esses resultados somente sejam válidos quando os índices de sucesso para

a angioplastia forem tão bons quanto os alcançados no contexto do ensaio clínico. Análises secundárias de bases de dados comunitários não mostraram benefício da angioplastia primária sobre a terapia trombolítica (8,9). Entretanto, é importante lembrar que estudos observacionais sobre tratamentos têm muitas limitações – a mais importante é o potencial para confundimento por diferenças nas características das pessoas tratadas e não tratadas. O viés e o confundimento são particularmente difíceis de avaliar utilizando bancos de dados secundários que não foram desenvolvidos para estudar a efetividade de um tratamento. Portanto, quando factíveis, ensaios clínicos randomizados que comparam tratamentos conduzidos na comunidade são uma abordagem melhor.

A análise de dados secundários, na maior parte das vezes, é a melhor forma de descrever como os tratamentos são usados na prática clínica. Embora ensaios clínicos possam demonstrar a eficácia de uma nova terapia, esse benefício somente ocorrerá se a terapia for de fato adotada pelos médicos na comunidade. A compreensão das **taxas de utilização** dos tratamentos e a consideração sobre suas **variações regionais** e uso em populações específicas (p. ex., idosos, minorias étnicas, camadas pobres da sociedade, mulheres) têm implicações relevantes para a saúde pública. Por exemplo, por meio de dados de acesso público de uma amostra aleatória de 5% dos beneficiários do Medicare, pesquisadores demonstraram importante variação regional na prevalência de diagnóstico de glaucoma, após ajuste para potenciais confundidores, sugerindo que há subdiagnóstico em certas regiões dos Estados Unidos (10).

Também é possível juntar duas ou mais bases de dados existentes para responder a uma questão de pesquisa, técnica essa conhecida como relacionamento ou linkagem de dados (*data linkage*). Investigadores interessados nos efeitos do serviço militar sobre a saúde usaram a convocação aleatória de 1970 a 1972, que envolveu 5,2 milhões de indivíduos do sexo masculino convocados aleatoriamente de acordo com a data de nascimento (primeira base de dados utilizada). Os investigadores fizeram o relacionamento desses dados com a mortalidade aferida a partir dos **registros de declarações de óbitos** (segunda fonte de dados). A variável preditora (*data de nascimento*) era uma variável substituta (*proxy*) para o serviço militar durante a guerra do Vietnã designada aleatoriamente. Homens que, devido à sua data de nascimento, haviam

sido designados aleatoriamente para o serviço militar nessa época tiveram uma mortalidade por suicídios e acidentes automobilísticos significativamente maior nos 10 anos subsequentes (11). O estudo teve um custo muito baixo, e, ainda assim, foi uma abordagem menos enviesada para examinar o efeito do serviço militar sobre causas específicas de morte do que outros estudos sobre esse mesmo tema com orçamentos muito maiores.

Quando dados individuais não estão disponíveis, podem ser usados dados agregados. **Dados agregados** incluem informações apenas para grupos de pessoas (p. ex., *mortalidade por câncer de colo uterino em cada um dos 50 Estados dos Estados Unidos*), e não para indivíduos. Com esses dados, só é possível medir associações correlacionando informações grupais sobre um determinado fator de risco (como *venda de cigarros por região*) com as respectivas taxas de um desfecho (câncer de pulmão por região). Estudos sobre associações com base em dados agregados são denominados **estudos ecológicos**.

A vantagem dos dados agregados é sua disponibilidade. Seu maior problema é sua suscetibilidade ao confundimento. Os grupos tendem a diferir em inúmeras formas, não apenas em relação à variável preditora de interesse. Como resultado, as associações observadas nos agregados não necessariamente se aplicam ao indivíduo. Por exemplo, a venda de cigarros pode ser maior em Estados com taxas elevadas de suicídio, mas os indivíduos que cometem suicídio podem não ser os que mais fumam. Essa situação é denominada **falácia ecológica**. Dados agregados encontram seu melhor uso no teste da plausibilidade de uma nova hipótese ou na formulação de novas hipóteses. Resultados interessantes podem, então, ser buscados em um novo estudo que use dados individuais.

Primeiros passos

Após escolher um tópico de pesquisa e familiarizar-se com a literatura na área (incluindo uma busca minuciosa da literatura e aconselhamento com um mentor sênior), o próximo passo é verificar se as questões de pesquisa escolhidas podem ser investigadas com um banco de dados já existente. A ajuda de um **pesquisador sênior** é de fundamental importância nesse processo. O investigador experiente já definiu suas áreas de interesse e se

mantém atualizado nessas áreas; além disso, conhece as bases de dados relevantes e os investigadores responsáveis por esses dados, tanto na sua instituição quanto fora dela. Essa pessoa pode ajudar a identificar e a obter acesso aos dados apropriados. Muitas vezes, a questão de pesquisa precisa sofrer algumas alterações (p. ex., modificando a definição das variáveis preditoras ou de desfecho) para se adequar aos dados disponíveis.

A melhor solução pode estar bem próxima do investigador, como uma **base de dados na própria instituição**. Por exemplo, um *fellow* da Universidade da Califórnia em São Francisco (UCSF), interessado no papel das lipoproteínas na doença coronariana, percebeu que uma das poucas intervenções que comprovadamente reduziam o nível de lipoproteína(a) era o estrogênio. Sabendo que o HERS, um importante ensaio clínico sobre terapia de reposição hormonal na prevenção de doença coronariana, era coordenado na UCSF, ele procurou os investigadores e manifestou seu interesse nessa questão. Como ninguém havia planejado estudar a relação entre essa lipoproteína, tratamento hormonal e eventos de doença coronariana, ele estudou a fundo a literatura sobre o assunto e delineou um plano de análise e de publicação. Após receber permissão da coordenação do estudo HERS, ele trabalhou com estatísticos, epidemiologistas e programadores do centro de coordenação para conduzir uma análise, que resultou em um artigo publicado em uma revista de prestígio (12).

Às vezes é possível estudar uma questão de pesquisa que tem pouco a ver com o estudo original. Por exemplo, um outro *fellow* da UCSF estava interessado no valor da repetição do exame de Papanicolau em mulheres com idade superior a 65 anos. Ele percebeu que a idade média das mulheres participantes do estudo HERS era de 67 anos, que elas deveriam ter um exame de Papanicolau normal antes de ingressarem no estudo e que o exame era repetido anualmente durante o seguimento. Acompanhando os resultados desse exame, ele conseguiu documentar que 110 esfregaços apresentaram resultado anormal dentre as 2.763 mulheres rastreadas no período de dois anos, e que apenas uma mulher mostrou ter um resultado anormal quando foi realizada a biópsia. Portanto, todos os resultados, exceto um, foram falso-positivos (13). Esse estudo teve forte influência sobre a recomendação seguinte do US Preventive Services

Task Force de que não se deve realizar o exame de Papanicolau em mulheres de baixo risco com idade superior a 65 anos com resultados prévios normais.

Muitas vezes, é necessária uma **empreitada maior**. Partindo de uma lista de variáveis preditoras e de desfecho cujas relações poderiam ajudar a responder à questão de pesquisa, o próximo passo é localizar bases de dados que contenham essas variáveis. Alguns estudos dispõem de portais na internet que fornecem acesso gratuito aos dados do estudo sem a necessidade de autorização prévia. Quando os dados não estão disponíveis on-line, telefonar ou enviar mensagens de e-mail para os autores de estudos anteriores ou para representantes do governo podem possibilitar o acesso a arquivos que contenham dados úteis. É fundamental vencer a ansiedade que pode resultar da ideia de contatar estranhos para solicitar ajuda. Em geral, outros pesquisadores se mostram bastante cooperativos, fornecendo dados ou sugerindo outros lugares onde esses dados podem ser obtidos.

Tendo localizado dados para responder à questão de pesquisa, o próximo desafio é obter a **permissão** para usá-los. Recomenda-se que se use correspondência em papel timbrado, bem como os títulos institucionais adequados, além de copiar o seu mentor como alguém que será reconhecido como especialista na área. É importante verificar se o seu mentor conhece os investigadores responsáveis pelo banco de dados, pois isso poderá facilitar o contato. Geralmente a forma mais eficaz de obter acesso aos dados é trabalhando em conjunto com um investigador envolvido no estudo de interesse, ou com um membro da equipe do estudo que esteja interessado no mesmo tópico de pesquisa que você. Esse investigador pode facilitar o acesso aos dados, assegurar que você compreenda a metodologia do estudo e como as variáveis foram medidas; além disso pode tornar-se um valioso colaborador. Bancos de dados de estudos multicêntricos e de ensaios clínicos geralmente têm procedimentos claros para o acesso aos dados, que incluem a exigência de uma proposta por escrito que precisa ser aprovada por um comitê de análise ou de publicações.

É importante ser bastante específico sobre **que informação** está sendo solicitada e confirmar a solicitação por escrito. Muitos estudos têm regras sobre como solicitar acesso aos dados que exigem que se especifiquem

quais dados estão sendo solicitados, como será feita a análise e o prazo para concluir o trabalho. É aconselhável ser conciso no texto e oferecer o pagamento de quaisquer custos envolvidos no preparo dos dados. Se a base de dados estiver sob o controle de um grupo de pesquisadores, pode-se sugerir uma colaboração. Isso incentiva o compartilhamento dos dados e engaja um coinvestigador familiarizado com eles. É recomendável que, desde o início, fique claramente definido em que consistirá essa colaboração, incluindo quem será o primeiro autor das publicações.

ESTUDOS SUPLEMENTARES

Em pesquisas com dados secundários, a maior parte dos dados necessários para responder a uma questão de pesquisa já estão disponíveis. Em um **estudo suplementar**, o investigador **acrescenta** uma ou mais **medidas** a um estudo já existente para responder a uma questão de pesquisa diferente. Por exemplo, no ensaio clínico HERS sobre o efeito da terapia de reposição hormonal no risco de eventos coronarianos em 2.763 mulheres idosas, um investigador acrescentou a medida da frequência e do grau de incontinência urinária. O acréscimo de apenas um breve questionário na visita seguinte gerou um ensaio clínico de grande porte sobre o efeito da terapia de reposição hormonal na incontinência urinária, com praticamente nenhum custo adicional em tempo ou recursos financeiros (14).

Os estudos suplementares têm muitas das **vantagens** da análise de dados secundários, com menos limitações. Têm baixo custo e são eficientes, e além disso o investigador pode delinear poucas medições-chave suplementares, especificamente para responder à questão de pesquisa. É possível acrescentar estudos suplementares a qualquer tipo de estudo, incluindo os transversais e de caso-controle, mas grandes estudos de coorte prospectivos e ensaios clínicos randomizados se adaptam particularmente bem a esse tipo de estudo.

Um problema de estudos suplementares é que as medições são mais informativas quando feitas antes do início do estudo, e é difícil para alguém de fora identificar estudos que estão em fase de planejamento. No entanto, mesmo quando uma variável não foi medida na linha de base, uma única medição durante o estudo ou no final dele pode produzir

informações bastante úteis. Por exemplo, acrescentando medições de função cognitiva no final do ensaio clínico HERS, os investigadores puderam comparar a função cognitiva em mulheres idosas tratadas com terapia de reposição hormonal, durante quatro anos, com a função cognitiva de mulheres tratadas com placebo (15).

Uma boa oportunidade para estudos suplementares é fornecida pelos **bancos de armazenamento de soro, DNA, imagens** e assim por diante, encontrados na maioria dos grandes ensaios clínicos e estudos de coorte. A oportunidade de propor novas medições nesses materiais armazenados pode ser uma abordagem extremamente custo-efetiva para responder a uma nova questão de pesquisa, sobretudo se for possível fazer essas medições em um subconjunto dos materiais, usando um delineamento de caso-controle aninhado ou de caso-coorte aninhado (Capítulo 8). No HERS, por exemplo, um estudo de caso-controle aninhado que realizou análises genéticas em amostras armazenadas demonstrou que o excesso do número de eventos tromboembólicos no grupo que recebeu terapia hormonal não era devido a uma interação com o fator V de Leiden (16).

Primeiros passos

Oportunidades para estudos suplementares devem sempre ser buscadas, especialmente por investigadores iniciantes com tempo e recursos limitados. Um bom ponto de partida é identificar estudos com questões de pesquisa que incluam a variável preditora ou a variável de desfecho de interesse. Por exemplo, um investigador interessado no efeito da perda de peso na dor associada à artrose do joelho poderia começar identificando estudos que incluem uma boa medida da dor relacionada à artrose (por meio de questionários validados) ou bancos de dados com registros sobre artroplastia que também haviam medido o peso anterior. Além disso, o investigador pode procurar por ensaios clínicos sobre intervenções (p. ex., *dieta, atividade física, mudança de comportamento, medicamentos*) na perda de peso. Esses estudos podem ser identificados buscando listas de pesquisas financiadas pelo governo federal (http://clinicaltrials.gov ou http://report.nih.gov), contatando indústrias farmacêuticas que produzem medicamentos para a perda de peso e conversando com especialistas em perda de peso que estejam familiarizados com os estudos em andamento.[3] Nesse caso, para elaborar um estudo suplementar, bastaria o investigador

acrescentar uma medição de sintomas de artrose em uma visita de seguimento dos sujeitos arrolados nesses estudos.

Após identificar um estudo que forneça uma boa oportunidade para medições suplementares, o próximo passo é obter o apoio dos investigadores do estudo. A maioria dos pesquisadores considera positivo acrescentar medições suplementares a um estudo já estabelecido, se elas abordarem uma questão importante e não interferirem substancialmente na condução do estudo principal. Por outro lado, esses mesmos pesquisadores teriam maior receio de acrescentar medições que exigiriam muito tempo do participante (p. ex., *teste de função cognitiva*), que seriam invasivos ou desagradáveis (*colonoscopia*) ou que acarretariam altos custos (*tomografia por emissão de pósitrons*).

Em geral, para fazer um estudo suplementar, é necessário ter permissão formal do investigador principal do estudo ou do comitê responsável. Grande parte dos estudos multicêntricos exigem uma solicitação por escrito. Frequentemente, o estudo suplementar proposto é revisado por um comitê, que pode aprovar, rejeitar ou sugerir alterações. Muitas medições suplementares requerem financiamento, e o investigador do estudo suplementar deve encontrar uma forma de cobrir esses custos. É claro que o custo de um estudo suplementar é muito menor do que o custo de conduzi-lo de forma independente. Estudos suplementares também são muito adequados para alguns tipos de financiamento do NIH que fornecem apenas um apoio pequeno para as medições e análises, porém um apoio substancial para o desenvolvimento da carreira (Capítulo 19). Alguns estudos maiores têm mecanismos próprios para financiar estudos suplementares, especialmente se a questão de pesquisa for importante e considerada relevante pela agência de financiamento.

São poucas as **desvantagens** dos estudos suplementares. Se o estudo for coletar dados dos participantes, é possível adicionar novas medições, mas as variáveis já aferidas geralmente não podem ser alteradas. Em alguns casos, pode haver problemas de ordem prática na obtenção da permissão formal dos investigadores ou da agência financiadora para a realização do estudo, no treinamento dos indivíduos que farão as medições e na obtenção separadamente do consentimento informado dos participantes. É preciso que antes do início do estudo essas questões fiquem esclarecidas, bem como a autoria dos artigos científicos resultantes do estudo

suplementar e as regras para o seu preparo e submissão.

REVISÕES SISTEMÁTICAS

Revisões sistemáticas identificam um conjunto de estudos já concluídos que abordam uma determinada questão de pesquisa e avaliam os resultados desses estudos para chegar a conclusões sobre um corpo de conhecimentos. Ao contrário de outras formas de revisar a literatura, uma revisão sistemática usa uma abordagem bem-definida para identificar todos os estudos relevantes, mostrar as características e os resultados dos estudos elegíveis e, quando apropriado, calcular uma estimativa-sumário dos resultados globais. Os **aspectos estatísticos** de uma revisão sistemática (cálculo das estimativas-sumário de efeito e variância, testes estatísticos de heterogeneidade e estimativas estatísticas de viés de publicação) são denominados **metanálise**.

Uma revisão sistemática pode ser uma excelente oportunidade para um investigador iniciante. Embora demande muito tempo e esforço, a revisão sistemática geralmente não exige muitos recursos. Completar uma boa revisão sistemática exige que o investigador adquira grande familiaridade com a literatura referente à questão de pesquisa. Para investigadores iniciantes, esse conhecimento detalhado da literatura publicada é de extremo valor. A publicação de uma boa revisão sistemática também pode dar a um investigador iniciante a posição de “perito” na questão de pesquisa. Além disso, os achados muitas vezes representam uma contribuição científica importante, resultante do maior poder estatístico dos estudos combinados ou então das peculiaridades nos achados de alguns estudos individuais, reveladas pela comparação com os achados dos outros estudos. Os achados de revisões sistemáticas são particularmente úteis no desenvolvimento de diretrizes para a **prática clínica**.

Os elementos de uma boa revisão sistemática são listados na Tabela 13.1. Uma boa fonte de informações sobre métodos para conduzir revisões sistemáticas de excelente qualidade pode ser encontrada no Cochrane Handbook for Systematic Reviews (http://handbook.cochrane.org). Assim como para outros estudos, os métodos para completar cada uma dessas etapas devem ser descritos em

um protocolo de pesquisa antes de ser iniciada a revisão sistemática.

TABELA 13.1 Elementos de uma boa revisão sistemática

1. Questão de pesquisa clara
2. Identificação abrangente e não enviesada dos estudos concluídos
3. Definição clara dos critérios de inclusão e exclusão
4. Extração uniforme e sem viés das características e dos achados de cada estudo
5. Apresentação clara e uniforme dos dados de estudos individuais
6. Cálculo de uma estimativa-sumário ponderada do efeito e do intervalo de confiança com base nos achados de todos os estudos elegíveis, quando apropriado
7. Avaliação da heterogeneidade dos achados dos estudos individuais
8. Avaliação do potencial para viés de publicação
9. Análises de subgrupo e de sensibilidade

Questão de pesquisa

Uma boa revisão sistemática requer uma questão de pesquisa bem-formulada e clara que atenda aos critérios **FINER** (Capítulo 2). A factibilidade depende em grande parte da existência prévia de um conjunto de estudos sobre a questão. A questão de pesquisa deve descrever a doença ou condição de interesse, a população e o contexto, a intervenção e o tratamento de comparação (para ensaios clínicos) e os desfechos de interesse. Por exemplo,

> "Em indivíduos admitidos em uma unidade de tratamento intensivo com síndrome coronariana aguda, o tratamento com aspirina mais heparina endovenosa reduz o risco de infarto do miocárdio e morte durante a hospitalização, quando comparado ao tratamento apenas com aspirina?"

Essa questão de pesquisa levou a uma metanálise que mostrou que adicionar aspirina à heparina melhora os desfechos e foi publicada em uma revista de prestígio (17), tendo importante impacto na prática clínica.

Identificando estudos já concluídos

Revisões sistemáticas baseiam-se em uma busca abrangente e sem viés de estudos já concluídos. A busca deve seguir uma estratégia bem definida estabelecida antes de se conhecerem os resultados dos estudos. Os procedimentos para identificar estudos para potencial inclusão na revisão sistemática e as fontes dos artigos devem ser explicitamente documentados antes de se iniciar a revisão. As buscas não devem se

limitar ao **MEDLINE**, que pode não listar as referências que não estejam na língua inglesa. Dependendo da questão de pesquisa, devem-se usar bases de dados eletrônicas como AIDS LINE, CANCERLIT e EMBASE, assim como revisões manuais da bibliografia de estudos relevantes publicados, revisões anteriores, avaliação da base de dados do Cochrane Collaboration e consultas a especialistas.[4] A estratégia de busca deve ser descrita claramente, de modo que outros investigadores possam replicá-la.

Critérios para inclusão e exclusão de estudos

O protocolo para a revisão sistemática deve fornecer uma boa justificativa para a inclusão e a exclusão de estudos, e esses **critérios devem ser estabelecidos *a priori*** (Tabela 13.2). Uma vez definidos esses critérios, cada estudo potencialmente elegível passa por uma revisão independente sobre elegibilidade feita por dois ou mais investigadores; as discordâncias entre esses investigadores devem ser resolvidas por outro revisor ou por consenso. Durante esse processo, é prudente cegar os revisores para a data, a revista, os autores e os resultados dos ensaios clínicos.

TABELA 13.2 Critérios para incluir ou excluir um estudo em uma metanálise

CRITÉRIOS	EXEMPLO: ÁCIDOS GRAXOS OMEGA-3 E EVENTOS CARDIOVASCULARES*
1. Período em que os estudos foram publicados	Estudos publicados antes de agosto de 2012
2. Delineamentos dos estudos	Ensaios clínicos randomizados sobre prevenção primária ou secundária de doença cardiovascular
3. População	Estudos com adultos randomizados para ácido graxo ômega-3 ou controle
4. Intervenção ou fator de risco	Administração de ômega-3, pela dieta ou por suplementos, em qualquer dose, por pelo menos um ano
5. Grupos-controle aceitáveis	Dieta ou suplemento sem ômega-3
6. Outras exigências relacionadas ao delineamento (p. ex., cegamento em ensaios clínicos ou controle para potenciais confundidores específicos em estudos observacionais)	Nenhuma
7. Desfechos aceitáveis	Mortalidade por todas as causas, morte cardíaca, morte súbita, infarto do miocárdio e acidente vascular encefálico
8. Perda máxima aceitável no seguimento	Não declarada
9. Duração mínima aceitável do seguimento	Não declarada

* Este exemplo de como esses critérios são usados foi extraído de uma metanálise publicada que mostrou ausência de efeito do ômega-3 na prevenção de eventos cardiovasculares.

Revisões sistemáticas publicadas devem **listar os estudos que foram considerados** para inclusão e o motivo específico para exclusão de um determinado estudo. Por exemplo, se 30 estudos potencialmente elegíveis foram identificados, esses 30 estudos devem ser listados com referência completa, devendo ser apresentados motivos para cada exclusão.

Coletando dados de estudos elegíveis

A extração de dados de cada estudo deve ser feita de maneira uniforme e livre de vieses. Em geral, isso é feito **independentemente por dois ou mais revisores** usando formulários pré-delineados (Tabela 13.3). Os formulários para extração de dados devem prever quaisquer dados que irão aparecer no texto, em tabelas ou em figuras, descrevendo os estudos incluídos na revisão sistemática ou em tabelas ou figuras sobre os desfechos. Quando houver discordância entre os dois revisores, um terceiro revisor pode tomar a decisão final, ou pode-se usar um processo de consenso. O processo de extração de dados de estudos para a revisão sistemática deve ser descrito de forma clara no manuscrito.

TABELA 13.3 Elementos que devem ser incluídos nos formulários de extração de dados para metanálises

1. Critérios de elegibilidade (o estudo atende a critérios de elegibilidade pré-estabelecidos?)
2. Características do delineamento (delineamento, grupo-controle, cegamento, controle para o cegamento, etc.)
3. Características e número de participantes em cada grupo de estudo (dados demográficos, gravidade da doença, etc.)
4. Intervenção (para ensaios clínicos) ou fatores de risco (para estudos observacionais)
   - Para intervenções – dose, duração do tratamento, etc.
   - Para estudos observacionais – tipo e nível do fator de risco, etc.
5. Desfecho principal, desfechos secundários e desfechos em subgrupos preestabelecidos
6. Elementos para permitir avaliação da qualidade dos estudos incluídos (randomização, cegamento, adesão, perda no seguimento, controle para confundimento, etc.)

Muitas vezes, os relatos publicados de alguns estudos potencialmente elegíveis para inclusão em uma revisão sistemática não incluem informações importantes, como características do delineamento, estimativas de risco e desvios-padrão. Nesses casos, é difícil dizer se determinadas características do delineamento, como cegamento, não foram implementadas ou se elas simplesmente não foram descritas na publicação. Pode ser possível calcular os riscos relativos e os ICs a partir

dos dados brutos apresentados em ensaios clínicos, mas em geral é inaceitável calcular estimativas de risco e ICs com base em dados brutos de estudos observacionais, pois não há informações suficientes para ajustar para possíveis confundidores. Todos os esforços devem ser tomados para que sejam obtidas dos autores informações importantes não incluídas na descrição publicada do estudo. No entanto, quando essas informações não podem ser calculadas ou obtidas, os achados do estudo são, em geral, excluídos.

Apresentando os achados de forma clara

Revisões sistemáticas geralmente incluem três tipos de informação. Em primeiro lugar, **características** importantes de cada estudo são apresentadas em tabelas. Essas tabelas em geral incluem características da população estudada, tamanho de amostra, número ou taxa de desfechos, duração do seguimento e métodos usados no estudo. Em segundo lugar, a revisão apresenta em uma tabela ou figura os **achados das análises** dos estudos individuais (riscos relativos, RCs, diferenças de riscos e ICs ou valores *P*). Por último, na ausência de heterogeneidade significativa (veja mais adiante), a revisão apresenta estimativas**-sumário e intervalos de confiança** com base nos achados de todos os estudos incluídos, bem como análises de sensibilidade e de subgrupos.

As estimativas-sumário de efeito representam os achados principais da revisão sistemática, mas devem ser apresentadas no contexto de todas as informações extraídas dos estudos individuais. As características e os achados de estudos individuais incluídos na revisão sistemática devem ser apresentados de forma clara em tabelas e figuras, de modo que o leitor possa formar sua própria opinião, não dependendo somente das estimativas estatísticas-sumário.

Metanálise: estatísticas para revisões sistemáticas

- **Estimativa-sumário de efeito e seu intervalo de confiança.** A partir do momento em que foram identificados todos os estudos já concluídos sobre a questão, escolhidos aqueles que atendam os critérios de inclusão e exclusão e extraídos os dados relevantes de cada estudo, geralmente se calcula uma **estimativa-sumário** (risco relativo-sumário, RCs-

sumário, diferença de riscos-sumário, etc.) e seu **IC**. O efeito-sumário é um efeito médio, ponderado pelo inverso da variância do desfecho de cada estudo. Métodos para calcular o efeito-sumário e seu IC são discutidos no Apêndice 13.1. Aqueles que não estiverem interessados nos detalhes de como calcular estimativas ponderadas médias a partir de múltiplos estudos devem pelo menos estar cientes de que diferentes abordagens fornecem diferentes resultados. Por exemplo, metanálises recentes sobre a eficácia do preservativo para prevenir a transmissão heterossexual do HIV forneceram estimativas-sumário variando de uma redução de 80 a 94% nas taxas de transmissão, embora essas estimativas se baseiem nos resultados de conjuntos praticamente idênticos de estudos (18,19).

- **Heterogeneidade.** Não é correto juntar os resultados de vários estudos se esses estudos diferirem de forma clinicamente significativa, como em aspectos relacionados à população, à intervenção, ao desfecho, à condição de controle, ao cegamento e assim por diante. Também não é correto juntar os achados se houver grandes diferenças entre os resultados dos estudos. Mesmo que as metodologias sejam aparentemente semelhantes, resultados discrepantes entre os estudos sugerem que deve ter havido alguma diferença importante. Essa variabilidade nos achados de estudos individuais é denominada heterogeneidade e diz-se que os achados são **heterogêneos**; quando há pouca variabilidade, diz-se que os achados são **homogêneos**.

Como decidir se as metodologias e os achados são suficientemente semelhantes para combinar os resultados em estimativas-sumário? Em primeiro lugar, pode-se revisar cada estudo e determinar se há diferenças substanciais no delineamento, nas populações estudadas, na intervenção ou no desfecho. Então, podem-se examinar os resultados dos estudos individualmente. Se alguns ensaios clínicos relatam um efeito significativamente benéfico de uma intervenção, e outros relatam danos consideráveis, fica claro que há heterogeneidade. Às vezes, porém, é mais difícil decidir se há heterogeneidade. Por exemplo, se um ensaio clínico relata que uma intervenção trouxe uma redução de 50% no risco, mas

outro ensaio clínico relata que a redução no risco foi de apenas 30%, há heterogeneidade? Abordagens estatísticas (testes de homogeneidade) foram desenvolvidas para ajudar a responder a essa questão (Apêndice 13.1), mas, na prática, a **avaliação da heterogeneidade** requer um julgamento crítico. Todas as revisões sistemáticas devem incluir uma discussão sobre a heterogeneidade e sobre seu efeito sobre as estimativas-sumário.

Avaliando o viés de publicação

O viés de publicação ocorre quando os estudos publicados não são representativos da totalidade de estudos realizados, pois é mais comum submeter e publicar resultados positivos do que resultados negativos. Há duas formas principais de lidar com o viés de publicação. Uma forma é **identificar estudos não publicados** e incluir seus resultados na estimativa-sumário. Esses resultados não publicados podem ser identificados indagando-se investigadores e revisando-se resumos, apresentações de encontros e teses de doutorado. Estudos não publicados podem ser identificados e os seus resultados incluídos junto com aqueles de estudos publicados na estimativa-sumário global. Também é possível fazer análises de sensibilidade para determinar se o acréscimo desses resultados não publicados poderia alterar de forma significativa a estimativa-sumário determinada a partir dos resultados publicados. Porém, incluir resultados não publicados em uma revisão sistemática apresenta inúmeros problemas. Muitas vezes é difícil identificar estudos não publicados, sendo ainda mais difícil extrair os dados necessários. Além disso, frequentemente as informações disponíveis são insuficientes para avaliar se o estudo atende os critérios de inclusão da revisão sistemática ou para avaliar a qualidade dos métodos empregados (que, por não terem sido submetidos ao rigor da revisão por pares, podem ser inferiores). Por esses motivos, não é comum incluir dados não publicados em metanálises.

Uma abordagem alternativa é **estimar** a magnitude do potencial viés de publicação e utilizar essa informação para ajustar as conclusões da revisão sistemática. O viés de publicação ocorre quando estudos não publicados apresentam achados diferentes daqueles de estudos publicados. É mais provável que os estudos não publicados sejam de pequeno tamanho

(estudos de grande porte são geralmente publicados, independentemente dos achados) e que não tenham encontrado associação entre o fator de risco ou a intervenção e o desfecho (estudos com resultados claramente positivos são, em geral, publicados, independentemente do tamanho). Se não houver viés de publicação, é improvável que haja associação entre os tamanhos dos estudos (ou a variância do desfecho) e seus achados. A força dessa associação pode ser medida usando o **tau de Kendall**, um coeficiente de correlação. Uma forte correlação entre os desfechos e os tamanhos de amostra sugere a possibilidade de viés de publicação. Na ausência de viés de publicação, se os tamanhos de amostra forem plotados contra os desfechos (p. ex., o logaritmo do risco relativo), o resultado deve apresentar curva em forma de sino ou de **funil**, com um ápice próximo da estimativa-sumário do efeito.

A curva em forma de funil na Figura 13.1A sugere que o viés de publicação foi pequeno, pois foram publicados estudos de pequeno porte com achados negativos e positivos. A curva na Figura 13.1B, por outro lado, sugere que houve viés de publicação, pois a distribuição foi truncada no canto que deveria conter estudos negativos de pequeno porte.

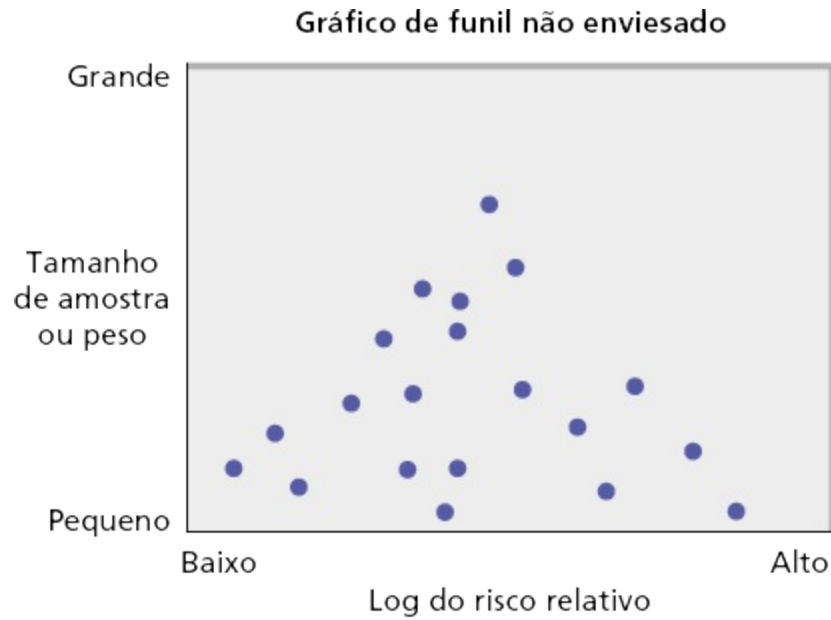
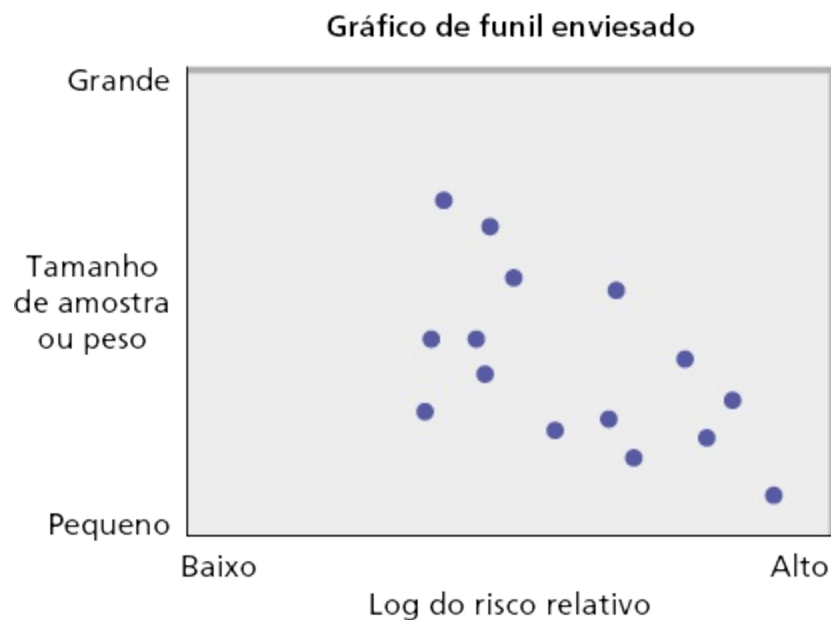



FIGURA 13.1 A: O gráfico de funil sugere que não há viés de publicação, pois há estudos com amostras grandes e pequenas, e são relatados riscos relativos baixos em estudos de menor tamanho. **B:** O gráfico de funil sugere que há viés de publicação, pois uma fração pequena dos estudos de menor tamanho relatam riscos relativos pequenos.

Quando for elevada a possibilidade de um viés nítido de publicação, a estimativa-sumário não deve ser calculada ou, pelo menos, deve ser

interpretada com cautela. Toda revisão sistemática deve conter uma discussão sobre potencial viés de publicação e seu efeito sobre as estimativas-sumário.

Análises de subgrupos e de sensibilidade

É possível fazer **análises de subgrupos** a partir de todos os estudos incluídos na revisão sistemática ou de um subconjunto deles. Por exemplo, em uma revisão sistemática sobre o efeito da terapia estrogênica pós-menopáusica no risco de câncer de endométrio, alguns estudos apresentaram os resultados de acordo com a duração da terapia estrogênica. A análise por subgrupos dos resultados desse subconjunto de estudos demonstrou que a duração maior da terapia estrogênica estava associada a maior risco de câncer (20).

As **análises de sensibilidade** mostram a "sensibilidade" dos achados da metanálise a decisões sobre aspectos do delineamento da revisão sistemática ou sobre a inclusão ou não de alguns dos estudos. Por exemplo, caso os autores decidam incluir na revisão sistemática estudos com delineamentos ou métodos levemente diferentes, os achados serão fortalecidos se os resultados-sumário forem similares, independentemente da inclusão de estudos questionáveis. Análises de sensibilidade devem ser feitas em revisões sistemáticas sempre que alguma decisão parecer questionável ou arbitrária no decorrer do estudo.

As metanálises podem aumentar o **poder** estatístico para responder a uma questão de pesquisa, mas têm a desvantagem de não incluírem dados de nível individual para permitir ajuste para potenciais confundidores ou para realizar análises de subgrupos individuais. Em algumas situações, pode ser possível obter os dados individuais a partir dos estudos e realizar **análises combinadas** (*pooled analyses*). Nesse caso, os dados combinados dos estudos individuais podem ser usados para ajustar para confundimento ou para avaliar efeitos de subgrupos, como seria feito em um estudo único maior. Por exemplo, o Early Breast Cancer Trialists Collaborative Group combinou os dados individuais de 123 ensaios clínicos randomizados para avaliar a eficácia de diferentes regimes de quimioterapia para câncer de mama inicial (21). Entretanto, geralmente é difícil obter dados individuais dos estudos relevantes e, na maioria das vezes, esses estudos não mediram as variáveis de forma suficientemente

semelhante para permitir que sejam combinadas em um único banco de dados.

Entra lixo, sai lixo (*garbage in*, *garbage out*)

A principal limitação da revisão sistemática é a possibilidade de produzir uma estimativa-sumário que aparenta ser bastante confiável quando, na verdade, se baseia nos resultados de estudos de baixa qualidade. Há diferentes abordagens para avaliar a qualidade dos diferentes delineamentos nas metanálises, mas o processo de avaliação da qualidade é complexo e problemático. Por isso, preferimos que os critérios de inclusão dos estudos sejam baseados em critérios rigorosos do bom delineamento de pesquisa. Se os estudos sumarizados na revisão sistemática forem de baixa qualidade, mesmo a mais criteriosa das análises não poderá impedir que a estimativa-sumário seja pouco confiável. Uma instância especial desse problema ocorre em revisões sistemáticas sobre dados observacionais. Se os resultados desses estudos não forem ajustados para potenciais confundidores, os resultados da metanálise também não estarão ajustados e irão potencialmente apresentar confundimento.

RESUMO

Este capítulo descreve três abordagens para fazer uso criativo de dados e materiais biológicos existentes, o que pode ser uma forma rápida e efetiva para pesquisadores iniciantes com recursos limitados adquirirem experiência valiosa e começarem a publicar.

Análise de dados secundários

1. Essa abordagem tem a **vantagem** de reduzir substancialmente o tempo e os custos da pesquisa e a **desvantagem** de permitir pouco ou nenhum controle sobre a população, o delineamento ou as aferições.
2. Algumas fontes de dados para análise secundária são **projetos de pesquisa já existentes, prontuários eletrônicos, bases de dados administrativos** e **bases de dados públicos**, como **registros de tumores, registros de declarações de óbito** e **inquéritos nacionais** como o **NHANES**.

3. Bancos de dados extensos de base comunitária são úteis para estudar a **efetividade** (efeitos de uma intervenção no mundo real em comunidades variadas); para avaliar as **taxas de utilização** e sua **variação regional**; e para detectar **eventos adversos raros.**
4. Estudos sobre associações que se utilizam de **dados agregados** são denominados **estudos ecológicos. E**sses estudos podem fornecer informações úteis, porém são sujeitos a vieses especiais denominados **falácias ecológicas**.

Estudo suplementar

1. O estudo suplementar é uma análise de dados secundários na qual o investigador realiza uma ou mais **novas aferições** para responder a uma nova questão de pesquisa com relativamente **poucos recursos e esforços**.
2. Boas oportunidades para estudos suplementares são encontradas em **estudos de coorte ou ensaios clínicos** que incluem a variável preditora ou de desfecho para a nova questão de pesquisa.
3. **Bancos de soro, DNA, imagens** e assim por diante oferecem a oportunidade para delineamentos de caso-controle aninhados.
4. Estudos de grande porte costumam ter **procedimentos** escritos, permitindo que investigadores (incluindo cientistas externos) proponham e conduzam análises de dados secundários e estudos suplementares.

Revisão sistemática

1. Uma boa revisão sistemática, assim como qualquer outro estudo, requer um **protocolo escrito** antes que o estudo possa começar. O protocolo deve incluir a **questão de pesquisa**, os métodos para **identificação de todos os estudos elegíveis**, os métodos para **extração de dados** dos estudos e os **métodos estatísticos**.
2. A parte estatística referente à combinação de estudos sobre um tópico, denominada **metanálise**, inclui a **estimativa-sumário de efeito e seu intervalo de confiança**, testes para avaliar a **heterogeneidade** e o **potencial viés de publicação** e análises de **subgrupos** e de **sensibilidade.**
3. As **características** e os **achados** de estudos individuais devem ser

apresentados claramente em tabelas e figuras, de modo que o leitor possa formar opiniões que não dependam unicamente das estatísticas-sumário.

4. Um desafio importante é avaliar a **qualidade** dos estudos em uma revisão sistemática, que pode influenciar fortemente os achados da revisão.

APÊNDICE 13
Métodos estatísticos para a metanálise

SUMÁRIOS DE EFEITO E INTERVALOS DE CONFIANÇA

O principal objetivo da metanálise é calcular um **sumário da estimativa de efeito** e o seu IC. Uma forma intuitiva de fazer isso é multiplicar cada desfecho do estudo, como o risco relativo (uma estimativa de efeito), pelo tamanho de amostra (um peso que reflete a precisão do risco relativo), somar os produtos e dividir o resultado pela soma dos pesos. Na prática, usa-se como peso para cada estudo o inverso da variância da estimativa do efeito de cada estudo individual (1/variância*i*). O inverso da variância é uma melhor estimativa da precisão da estimativa do efeito do que o tamanho de amostra, pois leva em consideração o número de desfechos e a sua distribuição. A média ponderada da estimativa do efeito é calculada multiplicando-se o peso de cada estudo (1/variância*i*) pelo logaritmo do risco relativo (ou qualquer outra estimativa de risco, como log da RC, diferença de risco, etc.); então, somam-se esses produtos e divide-se o resultado pela soma dos pesos. Estudos de pequeno porte geralmente apresentam grande variância (e um IC amplo em torno da estimativa de risco), e estudos de grande porte apresentam pequena variância (e um IC estreito em torno da estimativa de risco). Portanto, em uma metanálise, estudos de grande porte recebem um peso maior (1/pequena variância) e estudos de pequeno porte recebem peso menor (1/grande variância).

Para determinar se a estimativa-sumário do efeito é estatisticamente significativa, calcula-se a variabilidade da estimativa-sumário do efeito. Há várias fórmulas para calcular a variância das estimativas-sumário de risco (22,23). A maioria usa algo que se aproxima do inverso da soma dos pesos dos estudos individuais (1/soma dos pesos*i*). A variância da estimativa-sumário é usada para calcular o IC de 95% em torno da estimativa-sumário (±1,96 × variância$^{1/2}$).

MODELOS DE EFEITOS ALEATÓRIOS *VERSUS* MODELOS DE EFEITOS FIXOS

Há vários métodos estatísticos para calcular uma estimativa-sumário (22,23). A escolha do método estatístico geralmente depende do tipo de desfecho (risco relativo, RCs, diferença de risco, etc.). Além do método estatístico, deve-se também escolher entre usar um modelo de efeitos fixos ou de efeitos aleatórios. O **modelo de efeitos aleatórios** adiciona a variância ao efeito-sumário proporcionalmente à variabilidade dos resultados dos estudos individuais. As estimativas de efeito-sumário são geralmente semelhantes independentemente de se usar o modelo de efeitos fixos ou de efeitos aleatórios. Porém, a variância do efeito-sumário torna-se maior no modelo de efeitos aleatórios à medida que aumentam as diferenças entre os resultados dos estudos individuais. Da mesma forma, o IC em torno do efeito-sumário é correspondentemente maior, tornando a significância estatística menos provável. Muitas revistas exigem que os autores usem um modelo de efeitos aleatórios, por ser mais "conservador". As metanálises devem relatar de forma clara se usam o modelo de efeitos fixos ou aleatórios.

Simplesmente usar o modelo de efeitos aleatórios não elimina o problema da heterogeneidade. Se os estudos identificados por uma revisão sistemática forem claramente heterogêneos, não se deve calcular uma estimativa-sumário.

TESTES ESTATÍSTICOS DE HOMOGENEIDADE

Os testes de homogeneidade partem do pressuposto de que os achados dos ensaios clínicos são iguais (hipótese nula) e usam um teste estatístico (teste de homogeneidade) para verificar se os dados disponíveis (achados de estudos individuais) refutam essa hipótese. É comum usar o teste do qui quadrado para esse fim (22). Se os dados realmente apoiarem a hipótese nula ($P \geq 0{,}10$), o investigador aceita que os estudos são homogêneos. Caso contrário ($P < 0{,}10$), ele rejeita a hipótese nula e pressupõe que os achados são heterogêneos. Em outras palavras, conclui que há diferenças importantes nas populações estudadas, na natureza das variáveis preditora ou de desfecho ou nos resultados dos estudos.

Toda metanálise deve apresentar um teste de homogeneidade com um valor *P*. Esses testes não apresentam grande poder estatístico, e é difícil rejeitar a hipótese nula e provar a heterogeneidade quando o tamanho de

amostra (número de estudos usados) for pequeno. Por esse motivo, geralmente se usa como ponto de corte um valor *P* de 0,10, em vez de 0,05. Quando há uma importante heterogeneidade, não é correto juntar os resultados dos ensaios em uma única estimativa-sumário.

REFERÊNCIAS

1. Yaffe K, Barnes D, Nevitt M, et al. A prospective study of physical activity and cognitive decline in elderly women: women who walk. *Arch Intern Med* 2001;161:1703–1708.
2. Kerlikowske K, Miglioretti D, Buist D, et al. Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. *J Natl Cancer Inst.* 2007;99:1335–1339.
3. Canner PL. Mortality in CDP patients during a nine-year post-treatment period. *J Am Coll Cardiol* 1986;8:1243–1255.
4. Looker AC, Johnston CC Jr., Wahner HW, et al. Prevalence of low femoral bone density in older U.S. women from NHANES III. *J Bone Miner Res* 1995;10:796–802.
5. Looker AC, Melton LJ, Harris TB, et al. Prevalence and trends in low femur bone density among older US adults: NHANES 2005–2006 compared with NHANES III. *J Bone Miner Res* 2010;25:64–71.
6. Gurwitz JH, Gore JM, Goldberg RJ, et al. Risk for intracranial hemorrhage after tissue plasminogen activator treatment for acute myocardial infarction. Participants in the National Registry of Myocardial Infarction 2. *Ann Intern Med* 1998;129:597–604.
7. Weaver WD, Simes RJ, Betriu A, et al. Comparison of primary coronary angioplasty and intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review. *JAMA* 1997;278:2093–2098; published erratum appears in *JAMA* 1998;279:876.
8. Every NR, Parsons LS, Hlatky M, et al. A comparison of thrombolytic therapy with primary coronary angioplasty for acute myocardial infarction. Myocardial infarction triage and intervention investigators. *N Engl J Med* 1996;335:1253–1260.
9. Tiefenbrunn AJ, Chandra NC, French WJ, et al. Clinical experience with primary percutaneous transluminal coronary angioplasty compared with alteplase (recombinant tissue-type plasminogen activator) in patients with acute myocardial infarction: a report from the Second National Registry of Myocardial Infarction (NRMI–2). *J Am Coll Cardiol* 1998;31:1240–1245.
10. Cassard SD, Quigley HA, Gower EW, et al. Regional variations and trends in the prevalence of diagnosed glaucoma in the Medicare population. *Ophthalmology* 2012;119:1342–1351.
11. Hearst N, Newman TB, Hulley SB. Delayed effects of the military draft on mortality: a randomized natural experiment. *N Engl J Med* 1986;314:620–624.

12. Shlipak M, Simon J, Vittinghoff E, et al. Estrogen and progestin, lipoprotein (a), and the risk of recurrent coronary heart disease events after menopause. *JAMA* 2000;283:1845–1852.
13. Sawaya GF, Grady D, Kerlikowske K, et al. The positive predictive value of cervical smears in previously screened postmenopausal women: the Heart and Estrogen/Progestin Replacement Study (HERS). *Ann Intern Med* 2000;133:942–950.
14. Grady D, Brown J, Vittinghoff E, et al. Postmenopausal hormones and incontinence: the Heart and Estrogen/Progestin Replacement Study. *Obstet Gynecol* 2001;97:116–120.
15. Grady D, Yaffe K, Kristof M, et al. Effect of postmenopausal hormone therapy on cognitive function: the Heart and Estrogen/Progestin Replacement Study. *Am J Med* 2002;113:543–548.
16. Herrington DM, Vittinghoff E, Howard TD, et al. Factor V Leiden, hormone replacement therapy, and risk of venous thromboembolic events in women with coronary disease. *Arterioscler Thromb Vasc Biol* 2002;22:1012–1017.
17. Oler A, Whooley M, Oler J, et al. Heparin plus aspirin reduces the risk of myocardial infarction or death in patients with unstable angina. *JAMA* 1996;276:811–815.
18. Pinkerton SD, Abramson PR. Effectiveness of condoms in preventing HIV transmission. *Soc Sci Med* 1997;44:1303–1312.
19. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. *Cochrane Database Syst Rev* 2002;(1):CD003255.
20. Grady D, Gebretsadik T, Kerlikowske K, et al. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. *Obstet Gynecol* 1995;85:304–313.
21. Peto R, Davies C, Godwin J, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet* 2012;379:432–441.
22. Petitti DB. *Meta-analysis, decision analysis and cost effectiveness analysis: methods for quantitative synthesis in medicine*, 2nd ed. New York: Oxford University Press, 2000.
23. Cooper H, Hedges LV. *The handbook of research synthesis*. New York: Russell Sage Foundation, 1994.
24. Rizos EC, Ntzani EE, Bika E, et al. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events. *JAMA* 2012;308:1024–1033.

---

1 N. de R.T. No Brasil, estão disponíveis registros de câncer de base populacional em

http://www1.inca.gov.br/conteudo_view.asp?id=353 e, de base hospitalar, em http://www1.inca.gov.br/conteudo_view.asp?id=351.

2 N. de R.T. No Brasil, o estudo equivalente ao NHANES é a Pesquisa Nacional de Saúde (http://www.pns.icict.fiocruz.br/).

3 N. de R.T. No Brasil é possível pesquisar os estudos em andamento na Plataforma Brasil (http://aplicacao.saude.gov.br/plataformabrasil/visao/publico/pesquisarProjetoPesquisa.jsf).

4 N. de R.T. No Brasil, recomenda-se pesquisar também no portal das Bibliotecas Virtuais em Saúde (BVS), uma rede de bibliotecas temáticas e institucionais que abrange diversos temas e instituições relevantes na área da saúde (http://brasil.bvs.br). O site http://bvsalud.org permite fazer uma busca integrada em diversas bases de dados, incluindo MEDLINE, Scielo, Lilacs, Cochrane, entre outros. Permite também acesso ao texto integral das revisões sistemáticas da Cochrane.

SEÇÃO III

Implementação

CAPÍTULO 14

Abordando questões éticas

Bernard Lo e Deborah G. Grady

As pesquisas envolvendo seres humanos geram preocupações éticas, pois voluntários aceitam inconveniências e riscos com o objetivo de fazer avançar o conhecimento científico e beneficiar os outros. As pessoas que participam em pesquisas clínicas e ajudam a financiá-las precisam ter a confiança de que a pesquisa será conduzida de acordo com padrões éticos estritos.

Neste capítulo, começaremos abordando a **história** da supervisão regulatória das pesquisas e então revisaremos os **princípios éticos** e as **regulamentações** sobre pesquisas com seres humanos nos Estados Unidos[1], especialmente as exigências em relação à aprovação pelo **Comitê de Ética em Pesquisa (CEP)** e ao **consentimento informado**. Por fim, abordaremos questões relacionadas à **má-conduta científica**, ao **conflito de interesse** e às questões éticas em tipos específicos de pesquisa.

HISTÓRIA DAS REGULAMENTAÇÕES SOBRE A PESQUISA CLÍNICA

As recomendações e diretrizes atuais para a pesquisa clínica surgiram em resposta a abusos, incluindo pesquisas por médicos nazistas durante a

segunda guerra mundial, pesquisas realizadas em presidiários nos Estados Unidos, estudos com residentes de asilos e outras populações vulneráveis, bem como o Estudo Tuskegee (Caso 14.1).

CASO 14.1 Estudo Tuskegee (1)

Em 1932, agências do governo norte-americano iniciaram o estudo Tuskegee para documentar a história natural e os efeitos a longo prazo da sífilis não tratada. Os participantes da pesquisa eram homens negros, pobres e com baixa escolaridade de uma região rural do estado do Alabama. Eles receberam refeições, alguns cuidados médicos básicos e seguro funeral. Os pesquisadores mentiram para os participantes que eles estavam recebendo tratamento para sífilis, por exemplo, ao informar que as punções lombares feitas para fins de pesquisa eram “tratamentos especiais gratuitos”. Quando antibióticos para sífilis se tornaram disponíveis durante a segunda guerra mundial e foram posteriormente recomendados como medida de saúde pública, os pesquisadores tomaram medidas para evitar que os participantes da pesquisa recebessem o tratamento. Em resposta ao estudo Tuskegee, em 1974, o governo federal dos EUA lançou regulamentações sobre pesquisas em seres humanos, exigindo consentimento informado e revisão por CEPs para pesquisas em seres humanos financiadas pelo governo federal. Em 1997, o presidente Clinton pediu desculpas formais pelo estudo Tuskegee.

PRINCÍPIOS ÉTICOS

As falhas éticas do estudo Tuskegee e de diversos outros estudos inspiraram as regulamentações atuais para a proteção dos participantes de pesquisas. Três princípios éticos, que foram violados nesses estudos, foram articulados para nortear as pesquisas com seres humanos (2). Em primeiro lugar, reconhecendo que todas as pessoas têm o direito de tomar suas próprias decisões a respeito da participação na pesquisa, o princípio do **respeito à pessoa** exige que os investigadores obtenham consentimento informado dos participantes da pesquisa, permitam que eles abandonem a participação no estudo a qualquer momento e protejam

aqueles participantes com capacidade decisória reduzida.

Em segundo lugar, o princípio da **beneficência** exige que o conhecimento científico a ser adquirido a partir do estudo supere a inconveniência e o risco aos quais os participantes estão submetidos e que os riscos sejam minimizados. Os riscos incluem não somente danos físicos, mas também psicossociais, como quebra de confidencialidade, estigmatização e discriminação. Esses riscos podem ser minimizados, por exemplo, rastreando os possíveis participantes, a fim de excluir da pesquisa aqueles indivíduos suscetíveis a danos, assegurando a confidencialidade e monitorando os participantes para a ocorrência de efeitos adversos.

Em terceiro lugar, o princípio da **justiça** requer que os benefícios e o ônus da pesquisa sejam distribuídos de forma justa. Populações em desvantagem social ou vulneráveis, como aquelas com baixa renda, baixa escolaridade e acesso limitado a serviços de saúde, ou com capacidade decisória diminuída, não devem ser alvo especial da pesquisa se outras populações também forem adequadas para abordar essa questão de pesquisa. Estudar grupos vulneráveis primariamente devido à maior facilidade de acesso, à cooperação e ao seguimento pode significar usá-los de forma injusta.

O princípio da justiça também exige acesso equitativo aos benefícios da pesquisa. É importante ter isso em mente, pois tradicionalmente a pesquisa clínica tem sido considerada arriscada, e seus participantes vistos como cobaias que necessitam de proteção contra intervenções perigosas de pouco ou quase nenhum benefício. Entretanto, a pesquisa clínica tem se mostrado um veículo promotor de novas terapias para condições clínicas, como a infecção pelo vírus da imunodeficiência humana (HIV) e o câncer. Pacientes que buscam medicamentos novos promissores para doenças letais querem maior acesso à pesquisa clínica, e não maior proteção contra ela, e esse acesso precisa estar disponível independentemente de renda, acesso a seguro de saúde ou escolaridade. Historicamente, crianças, mulheres e membros de grupos étnicos minoritários têm sido sub-representados na pesquisa clínica, o que resulta em uma base de evidências mais fraca e, potencialmente, a um cuidado clínico inferior. Pesquisadores clínicos financiados pelo NIH devem assegurar representação adequada de crianças, mulheres e minorias

étnicas em seus estudos ou, então, justificar o motivo pelo qual tais grupos estão sendo sub representados.

REGULAMENTAÇÕES FEDERAIS PARA PESQUISAS EM SERES HUMANOS NOS ESTADOS UNIDOS

A **regulamentação federal norte-americana** se aplica a todas as pesquisas financiadas pelo governo federal daquele país e às pesquisas submetidas ao Food and Drug Administration (FDA) para apoiar a liberação de um novo medicamento ou dispositivo. Além disso, as universidades exigem que as pesquisas com participantes humanos conduzidas por docentes ou funcionários obedeçam às diretrizes básicas referentes ao consentimento informado e à aprovação pelo CEP, mesmo quando financiadas por fundos privados ou conduzidas em outro local. Embora na regulamentação seja utilizada a expressão “sujeitos de pesquisa”, o termo “**participantes**” é preferido por alguns autores, pois enfatiza que as pessoas são participantes ativas na pesquisa, em vez de sujeitos submetidos a um experimento.

Essas regulamentações contêm diversas **definições** que precisam ser compreendidas:

- **Pesquisa** é uma “uma investigação sistemática delineada para desenvolver ou contribuir para o desenvolvimento de conhecimentos generalizáveis” (3). O atendimento médico não testado formalmente, que se direciona para o benefício individual do paciente e não para a publicação de dados, não é considerado pesquisa. Alguns projetos de melhoria de qualidade poderiam ser considerados pesquisa, embora a maioria deles se enquadrem nos critérios para isenção,[2] como será discutido adiante.

- **Sujeitos de pesquisa** são pessoas vivas sobre as quais um investigador obtém “dados por meio de intervenções ou de interação com o indivíduo” ou “informações privadas identificáveis”.

- **Informações privadas** compreendem (1) informações que uma pessoa esperaria normalmente não estarem sendo observadas ou registradas e (2) informações fornecidas para fins específicos e que “o indivíduo

pode esperar que não sejam tornadas públicas (p. ex., dados do prontuário médico)". A informação é dita identificável quando "a identidade do indivíduo é ou pode ser facilmente determinada pelo investigador".

- Dados de pesquisa codificados **não são identificáveis** se a lista que relaciona os dados aos respectivos participantes for destruída antes do início do estudo ou se os investigadores não tiverem acesso a essa lista.

As Regulamentações Federais sobre Proteção de Sujeitos Humanos (*Federal Regulations on the Protection of Human Subjects*) estão disponíveis na página na internet do *Office for Human Research Protections* (3).[3] Os pesquisadores que tiverem dúvidas sobre as normas vigentes devem consultar seu CEP. As regulamentações federais oferecem dois tipos de proteção aos sujeitos de pesquisas – a aprovação pelo CEP e o consentimento informado.

Aprovação pelo Comitê de Ética em Pesquisa

As pesquisas que envolvam participantes humanos devem ser submetidas à apreciação de um **CEP** (IRB, *Institutional Review Board*).[4] A missão do comitê é assegurar que a pesquisa seja eticamente aceitável e que os direitos e bem-estar de seus participantes sejam protegidos. Embora a maioria dos membros do comitê seja composta por pesquisadores, sua constituição deverá incluir também membros da comunidade e pessoas com conhecimento em questões éticas e legais no campo da pesquisa.

Ao aprovar uma pesquisa, o comitê deve determinar que (3):

- Os **riscos** aos participantes sejam **minimizados**.
- Os **riscos se justifiquem** pelos benefícios antecipados e pela importância dos conhecimentos a serem adquiridos com os resultados.
- A **seleção** dos participantes seja **equitativa**.
- O **consentimento informado** seja obtido dos participantes ou de seus representantes legalmente autorizados.
- A **confidencialidade** seja mantida.

O sistema de CEPs é descentralizado. Cada comitê local implementa a legislação federal usando diretrizes e formulários próprios[5], e não há apelação a uma instância superior. Como consequência, um estudo multicêntrico pode ser aprovado pelo comitê de uma instituição e não pelo comitê de outra. Problemas daí decorrentes podem ser resolvidos por meio de discussões ou por alterações do protocolo.

Os CEPs e as regulamentações federais têm sido criticados por várias razões (4,5). Eles podem colocar ênfase excessiva nos formulários de consentimento, deixar a desejar no que diz respeito à avaliação do delineamento da pesquisa e não considerar adequadamente o mérito científico do estudo. Embora os comitês sejam responsáveis pela revisão de quaisquer alterações nos protocolos e pelo monitoramento de eventos adversos, eles não averiguam se a pesquisa foi realmente conduzida de acordo com o que foi estabelecido nos protocolos. Por carecerem de recursos e pessoal qualificado, muitos CEPs falham no cumprimento de sua missão precípua, que é a proteção aos participantes da pesquisa. Por essas razões, as regulamentações federais e a aprovação pelo comitê de ética devem ser consideradas como um padrão mínimo de ética em pesquisa. O **juízo e o caráter do investigador** são os elementos mais importantes para garantir que a pesquisa seja eticamente aceitável.

Exceções à revisão completa pelo Comitê de Ética em Pesquisa

- A maior parte das pesquisas que utilizam inquéritos populacionais e entrevistas, assim como as análises secundárias de dados não identificados oriundos de registros e amostras **preexistentes** podem ser isentos da revisão pelo CEP (Tabela 14.1). A justificativa ética para essas isenções é que a pesquisa envolve baixo risco, quase todas as pessoas iriam consentir e obter o consentimento de cada participante tornaria o estudo proibitivamente caro ou difícil. Muitos CEPs, entretanto, exigem que os pesquisadores submetam alguma informação sobre o projeto, para verificar se ele se qualifica para a isenção.

- O CEP pode permitir que alguns estudos que envolvem riscos mínimos sejam submetidos a uma **revisão expedita**[6] por um único revisor, em vez de uma avaliação pelo comitê inteiro (Tabela 14.2). O *site* na

internet do *Office for Human Research Protections* lista os tipos de pesquisa elegíveis para revisão expedita (6). O conceito de **risco mínimo aos participantes** desempenha papel importante junto às leis federais, como indicado na Tabela 14.2. Risco mínimo é definido como aquele "geralmente encontrado na vida cotidiana ou durante a realização de avaliações físicas ou psicológicas de rotina". Deve-se considerar tanto a magnitude quanto a probabilidade do risco. O CEP deve sempre julgar se um determinado projeto pode ser considerado de risco mínimo.

TABELA 14.1 Pesquisas isentas das normas federais norte-americanas

1. Inquéritos, entrevistas ou observações de comportamento público, exceto:
   - Quando haja risco de identificação do sujeito, e
   - Quando a divulgação das respostas dos sujeitos do estudo possa colocá-los em risco perante a lei ou denegrir sua reputação, situação financeira ou empregabilidade. Por exemplo, questionários sobre temas como dependência química, depressão, comportamentos de risco para HIV ou imigração ilegal não estão isentos.
2. Estudos que utilizam registros, dados ou amostras já existentes, desde que:
   - Eles estejam disponíveis ao público (p.ex, bancos de dados disponibilizados por agências estaduais ou federais) OU
   - As informações tenham sido registradas pelo investigador de maneira que os sujeitos da pesquisa não possam ser identificados.
3. Pesquisas sobre práticas regulares de ensino.

TABELA 14.2 Tipos de pesquisa que podem ser submetidos à revisão expedita pelo CEP

1. Certos procedimentos que envolvem risco mínimo, incluindo:
   - Coleta de sangue venoso, saliva ou escarro, bem como swabs de pele ou de mucosas.
   - Coleta de amostras por meio de procedimentos não-invasivos empregados rotineiramente na prática clínica tais como eletrocardiogramas e ressonância magnética. Entretanto, os procedimentos que utilizam raio X, que expõem os pacientes à radiação, devem ser revisados pelo comitê.
   - Pesquisas envolvendo dados, registros ou amostras previamente coletados ou que serão coletados para propósitos clínicos.
   - Pesquisas que utilizam inquéritos ou entrevistas que não estejam isentos da revisão pelo comitê de ética.
2. Alterações menores em protocolos de pesquisa previamente aprovados.
3. Renovação de aprovação pelo CEP para estudos que já tenham sido concluídos exceto pela análise de dados ou pelo seguimento a longo prazo.

Consentimento informado e voluntário

Os investigadores devem obter o consentimento informado e voluntário dos participantes da pesquisa.

Divulgação de informações aos participantes

As regulamentações federais exigem que os investigadores discutam com os potenciais participantes diversos tópicos, incluindo:

- **A natureza do projeto de pesquisa.** O potencial participante deve ser informado explicitamente de que uma pesquisa está sendo realizada, do objetivo da pesquisa e quem está sendo recrutado. Não é necessário declarar a hipótese específica do estudo.

- **Os procedimentos do estudo.** Os participantes devem saber o que se espera deles no projeto de pesquisa. Em termos práticos, devem ser informados sobre quanto tempo será exigido e com que frequência. Procedimentos que não fazem parte dos cuidados clínicos usuais devem ser identificados como tal. Se o estudo envolver cegamento ou randomização, esses conceitos devem ser explicados em uma linguagem que o participante possa compreender. Em pesquisas que envolvam entrevistas ou questionários, os participantes devem ser informados sobre os assuntos que serão abordados.

- **Os riscos e potenciais benefícios e as alternativas à participação no estudo.** Os riscos e benefícios médicos, psicossociais e econômicos devem ser descritos em linguagem leiga. Além disso, os potenciais participantes devem ser informados sobre as alternativas à participação, por exemplo, se a intervenção oferecida por um ensaio clínico estará disponível fora do estudo. Uma preocupação que vem sendo relatada é que muitas vezes as informações oferecidas aos participantes dão pouca ênfase aos riscos e hiperdimensionam os benefícios (7). Por exemplo, pesquisas sobre novos medicamentos são às vezes descritas como oferecendo benefícios aos participantes. No entanto, a maioria das novas intervenções promissoras, apesar dos resultados preliminares encorajadores, não mostram vantagens significativas em relação ao tratamento-padrão. Muitas vezes os participantes ficam com a falsa percepção de que a intervenção sob estudo deve oferecer benefícios pessoais a eles (8). Os investigadores devem deixar claro que não se sabe se o medicamento ou intervenção sob estudo é mais eficaz do que o tratamento-padrão, e que novas medicações promissoras podem causar danos graves à sua saúde.

Formulários de consentimento

Os formulários para consentimento por escrito devem documentar a ocorrência do processo de consentimento informado – isto é, a discussão entre o investigador e o participante. O formulário de consentimento deve conter todas as informações necessárias discutidas na seção anterior. Uma alternativa é usar um formulário breve, em que conste que os elementos necessários do consentimento informado foram apresentados oralmente. Se for usada a forma breve do formulário, deve haver uma pessoa que testemunhe a apresentação oral e assine o formulário junto com o participante.[7]

Os CEPs geralmente dispõem de modelos de formulários que apresentam o formato que eles preferem que seja utilizado. Os CEPs podem exigir que mais informações sejam divulgadas do que o requerido pelas normas federais.[8]

Compreensão pelos participantes das informações divulgadas

Em termos éticos, a questão mais importante relacionada ao consentimento não é quais informações o pesquisador divulga, mas se os participantes compreendem os riscos e benefícios do projeto. Os participantes muitas vezes têm concepções equivocadas sobre os objetivos da pesquisa e sobre os procedimentos e os riscos do protocolo específico do estudo (9). Nas conversas com os participantes e nos formulários de consentimento, deve-se evitar jargão técnico e frases complicadas. Os CEPs têm sido criticados por focarem sua atenção excessivamente nos formulários de consentimento, em vez de se certificarem de que os participantes compreendem as informações mais importantes (9). Estratégias para aumentar a **compreensão** dos participantes incluem ter um membro do estudo ou um educador neutro com mais tempo disponível para conversar individualmente com os participantes, simplificar os formulários de consentimento, usar um formato de pergunta e resposta, fornecer informações ao longo de várias consultas e usar fitas informativas de áudio ou de vídeo (10). Em pesquisas que envolvam riscos substanciais ou que sejam controversas, deve-se considerar avaliar a compreensão dos participantes e documentar quando eles conseguem responder corretamente a questões relacionadas aos aspectos mais importantes da pesquisa (11, 12).

Natureza voluntária do consentimento

O consentimento eticamente válido deve ser voluntário, assim como informado. Deve-se minimizar a possibilidade de os participantes serem coagidos a participar, ou de serem influenciados de forma indevida. Exemplos de influência indevida são pagamentos excessivos aos participantes ou incluir estudantes como participantes da pesquisa. Uma influência indevida é eticamente problemática se levar os participantes a minimizar de forma significativa os riscos de uma pesquisa ou comprometer gravemente sua capacidade de se recusar a participar. Os participantes devem compreender que a recusa em participar no estudo não irá comprometer seu atendimento médico e que eles podem se desligar do projeto a qualquer momento.

Exceções ao consentimento informado[9]

Alguns estudos com grande importância científica se tornariam difíceis ou impossíveis de realizar se o consentimento informado fosse exigido de cada participante.

Pesquisas com sobras de materiais biológicos e de dados não identificados

CASO 14.2 Pesquisa com amostras de sangue neonatal

Logo após o nascimento, é feita punção com uma lanceta no calcanhar do recém-nascido para coletar sangue em papel-filtro com o objetivo de rastrear doenças genéticas (“teste do pezinho”). Na maior parte dos Estados Unidos, não é necessária permissão dos pais para esse rastreamento obrigatório, portanto as amostras representam toda a população de recém-nascidos. O que sobra de sangue após o rastreamento clínico tem sido valioso para pesquisas sobre causas genéticas de malformações congênitas e de parto pré-termo, exposições ambientais durante a gestação e interações gene-ambiente.

O consentimento informado e a revisão pelo CEP não são necessários quando se utilizam amostras biológicas não identificadas (Tabela 14.1), mas muitos CEPs ainda exigem que os investigadores comuniquem quando estão realizando essas pesquisas. Quando uma pesquisa original

desse tipo é publicada, muitas revistas exigem que os autores declarem que o CEP aprovou o protocolo ou determinou que a revisão não era necessária.

Dispensa da necessidade de consentimento informado

Alguns projetos de pesquisa muito relevantes exigem a utilização de informações e amostras existentes e identificadas. Esses estudos não estão isentos de revisão pelo CEP, mas podem se qualificar para dispensa do consentimento informado.

CASO 14.2 Pesquisa com amostras de sangue neonatal *(continuação)*

Uma equipe de pesquisa gostaria de utilizar amostras de sangue neonatal identificadas para estudar a associação entre exposição materna a determinadas substâncias químicas e baixo peso ao nascer, prematuridade e morte perinatal. É possível relacionar as amostras identificadas a declarações de nascidos vivos, a declarações de óbito e a registros hospitalares. Devido ao grande número de crianças que precisariam ser estudadas para alcançar poder estatístico suficiente para detectar associações, não seria factível obter permissão dos pais ou guardiões.

Conforme a legislação federal norte-americana, os CEPs podem dispensar um estudo da obrigatoriedade de obter consentimento informado se todas as condições na Tabela 14.3 estiverem presentes. A maioria dos CEPs provavelmente dispensaria o consentimento informado no caso do estudo sobre exposições ambientais em gestantes proposto no exemplo anterior.

TABELA 14.3 Pesquisas nas quais é possível obter dispensa do consentimento informado

1. A pesquisa não oferece mais que o risco mínimo aos participantes; e
2. A liberação ou alteração não afetaria os direitos e o bem-estar dos participantes; e
3. A pesquisa tornar-se-ia, de outra forma, impraticável; e
4. Quando apropriado, os sujeitos receberão informações adicionais pertinentes após sua participação no estudo. Essa concessão permite despistar o motivo da pesquisa, por exemplo, quando revelá-lo comprometeria a validade do estudo.

Justificativa para isenção de consentimento informado

Algumas pesquisas com elevada importância científica apresentam riscos tão baixos que obter o consentimento seria trabalhoso demais e ao mesmo tempo faria pouco para proteger os participantes. Todos os pacientes já se beneficiaram de conhecimentos gerados a partir de estudos que utilizaram registros médicos e amostras biológicas previamente existentes. A justiça, no sentido de reciprocidade, sugere que as pessoas que se beneficiam dessas pesquisas aceitariam participar de estudos semelhantes de baixíssimo risco que poderiam beneficiar outras pessoas.

Objeções à isenção de consentimento informado

Embora a legislação federal norte-americana permita que amostras de sangue neonatal não identificadas sejam usadas para fins de pesquisa sem aprovação dos pais, há forte oposição pública a isso.

CASO 14.2 Pesquisa com amostras de sangue neonatal *(continuação)*

Pais em diversos estados se opõem ao armazenamento de amostras para pesquisas não especificadas sem sua autorização prévia ou sem a oportunidade de se retirar da pesquisa. Isso levou a ações judiciais em dois Estados norte-americanos. Os autores das ações não contestaram a coleta de sangue para rastreamento neonatal, porém argumentaram que mesmo o fato de as amostras não estarem identificadas não foi suficiente para conter sua inquietação em relação à perda de privacidade e de autonomia.

Uma vez que tais objeções podem comprometer a captação de crianças para o rastreamento neonatal, vários Estados estão começando a oferecer aos pais a oportunidade de optar por não participar dessas pesquisas. Tal atenção às inquietações dos pais pode ir além do que as normas federais exigem. Portanto, o que é legalmente permitido na pesquisa pode nem sempre ser eticamente aceitável, em especial no caso de pesquisas sobre temas sensíveis.

Participantes com capacidade decisória reduzida

Quando os participantes não são capazes de fornecer consentimento informado, a permissão para participar do estudo deve ser obtida de seu

representante legal (pai, mãe ou pessoa com a guarda legal, no caso de crianças pequenas). Além disso, o protocolo da pesquisa deve ser sujeito a uma avaliação mais rigorosa, para averiguar se a questão de pesquisa não poderia ser estudada em uma população capaz de fornecer o consentimento informado.

Minimizando riscos

Os pesquisadores devem antecipar os riscos que poderão ocorrer nos projetos de pesquisa e reduzi-los, por exemplo, por meio da identificação e exclusão das pessoas mais suscetíveis a efeitos adversos, do monitoramento desses eventos adversos e da introdução de métodos de aferição menos invasivos. Um aspecto importante dessa minimização de riscos é a manutenção da confidencialidade dos participantes.

Confidencialidade

Quebras de confidencialidade podem causar estigma ou discriminação, especialmente se o estudo estiver abordando tópicos sensíveis, como atitudes ou práticas sexuais, uso de álcool ou drogas, condutas ilegais e doenças psiquiátricas. Estratégias para proteger a confidencialidade incluem codificar os dados da pesquisa, proteger ou destruir dados que identifiquem os participantes e limitar o número de pessoas com acesso aos identificadores. No entanto, não se deve fazer promessas incondicionais de confidencialidade. A confidencialidade pode ser revogada se os registros da pesquisa forem submetidos à auditoria, se houver determinação legal, ou se forem identificadas condições clínicas que legalmente exijam notificação, como maus-tratos a crianças, certas doenças infecciosas e graves ameaças de violência. Em projetos que preveem situações como essas, o protocolo deve especificar como os membros da pesquisa devem proceder nesses casos, e os participantes devem ser informados sobre esses procedimentos.

Nos Estados Unidos, o pesquisador pode prevenir intimações judiciais obtendo **certificados de confidencialidade**[10] do Public Health Service (13), que permite reter dados de pesquisa identificáveis no caso de determinação ou ordem judicial para revelá-los. Entretanto, esses certificados ainda não foram amplamente testados no contexto de decisões judiciais, não se aplicam a auditorias por agências de financiamento ou

pelo FDA e não liberam o pesquisador de revelar voluntariamente informações relacionadas a maus-tratos a crianças e idosos, violência doméstica ou doenças transmissíveis de notificação compulsória. A pesquisa não precisa ser financiada por órgão federal para receber um certificado de confidencialidade.

Regulamentação sobre privacidade em saúde da HIPAA[11]

A regulamentação federal sobre privacidade em saúde nos Estados Unidos (conhecida pela sigla **HIPAA**, do inglês *Health Insurance Portability and Accountability Act*) protege informações que possam permitir a identificação de um indivíduo, denominadas **informações protegidas de saúde**. De acordo com essa norma, os indivíduos devem assinar uma autorização para permitir que o provedor de cuidados de saúde use ou divulgue informações protegidas de saúde em um projeto de pesquisa (14). O formulário de autorização da HIPAA não substitui o formulário de consentimento informado exigido pelo CEP. Os investigadores devem obter autorização para cada uso de informações protegidas de saúde para fins de pesquisa; não é permitido consentimento geral para pesquisas futuras. A autorização não é necessária caso os dados não sejam identificáveis e em algumas outras situações. Os pesquisadores devem contatar seu CEP em caso de dúvidas sobre essa norma de privacidade e sobre como ela difere das normais federais para proteção de sujeitos humanos.

PARTICIPANTES DE PESQUISA QUE REQUEREM PROTEÇÕES ADICIONAIS

Alguns participantes podem apresentar “maior risco de serem usados de forma eticamente inapropriada em pesquisa”, devido à maior dificuldade para fornecer o consentimento voluntário e informado ou à maior suscetibilidade a eventos adversos (15).

Tipos de vulnerabilidade

A identificação de diferentes tipos de vulnerabilidade permite aos investigadores adotar salvaguardas que se apropriem ao tipo específico de vulnerabilidade dos participantes do estudo.

Limitações cognitivas ou comunicativas

Indivíduos com limitação da função cognitiva ou da capacidade de comunicação podem ter dificuldade para compreender as informações sobre um estudo e para pesar os seus riscos contra seus benefícios.

Diferenças de poder

Indivíduos institucionalizados, como presidiários ou idosos que moram em asilos, podem se sentir pressionados a participar de pesquisas e a se submeterem a pessoas que controlam sua rotina diária. Assim, esses indivíduos podem achar que a recusa em participar da pesquisa poderá levar a retaliações por parte das autoridades da instituição ou comprometer outros aspectos de sua rotina diária.

Se o investigador de um estudo também for o **médico responsável** pelo atendimento de um participante, este poderá hesitar em se recusar a participar, com receio de que o médico terá menos interesse no seu cuidado. Da mesma forma, alunos e estagiários podem se sentir pressionados a participar de pesquisas conduzidas por seus instrutores ou supervisores.

Desvantagens econômicas e sociais

Pessoas em situação socioeconômica desfavorável ou com menor acesso a cuidados de saúde podem ingressar em um estudo para obter pagamento ou cuidados médicos, mesmo que, se tivessem maior renda, considerariam os riscos como inaceitáveis. Participantes com baixa escolaridade ou com poucos conhecimentos médicos podem não compreender informações sobre o estudo ou estar mais suscetíveis a influências de outras pessoas.

Proteções para participantes vulneráveis

As normas federais norte-americanas para pesquisas com participantes vulneráveis podem ser encontradas no *site* do Office for Human Research Protections (3).

Pesquisas com crianças

Os investigadores devem obter permissão dos pais e da própria criança caso esta seja capaz de manifestar seu consentimento. Pesquisas em crianças envolvendo riscos além dos mínimos são permitidas apenas nas

seguintes situações:

- Quando há perspectivas de benefícios diretos para a criança, ou
- Se ela ultrapassar apenas ligeiramente o risco mínimo e se apresentar boa possibilidade de originar conhecimentos generalizáveis importantes sobre a doença ou condição clínica da criança.

Pesquisas com presidiários

Os presidiários podem não se sentir à vontade para declinar a participação na pesquisa e podem ser indevidamente influenciados por dinheiro, quebras na rotina da prisão ou por promessas de liberdade. As regulamentações federais limitam os tipos de pesquisas com presidiários e exigem uma apreciação mais minuciosa pelo CEP e aprovação pelo *Department of Health and Human Services*.

Pesquisas com gestantes, fetos e embriões

Pesquisas que não oferecem perspectiva de benefício direto ao feto são permitidas apenas "se o objetivo da pesquisa for desenvolver conhecimentos biomédicos importantes que não podem ser obtidos de outra forma". Pesquisas que oferecem perspectiva de benefício apenas para o feto requerem consentimento informado do pai e da mãe, embora as pesquisas que oferecem perspectiva de benefício direto para crianças exijam apenas a permissão de um dos progenitores. Essas restrições têm sido criticadas por limitarem pesquisas que podem fortalecer a base de evidências para o cuidado clínico de gestantes e fetos.

RESPONSABILIDADES DOS INVESTIGADORES

Diversos casos graves de más-condutas em pesquisa continuam sendo relatados ainda hoje.

CASO 14.3 Efeitos adversos cardíacos do rofecoxibe

Em 2000, foram publicados os resultados do ensaio clínico randomizado VIGOR, que comparou um novo anti-inflamatório não esteroide seletivo para a COX-2, o rofecoxibe, com um anti-inflamatório mais antigo, não seletivo, o naproxeno (16). O estudo era

financiado pelo fabricante do rofecoxibe. O novo medicamento causou um número significativamente menor de complicações gastrintestinais do que o naproxeno (2,1 vs. 4,5 por 100 pacientes-anos), ao mesmo tempo em que tinha uma eficácia semelhante para a dor articular. O grupo que recebeu rofecoxibe também tinha mais casos de infarto (0,4% vs. 0,1%). Após essa publicação, o rofecoxibe passou a ser amplamente prescrito, com vendas excedendo 2,5 bilhões de dólares anuais. Antes da publicação do artigo, três casos adicionais de infarto no grupo do rofecoxibe foram relatados ao FDA, mas não aos autores do artigo que eram ligados à universidade ou à revista. Dois autores que eram empregados do fabricante do medicamento sabiam desses casos adicionais. A revista que publicou os resultados do estudo VIGOR posteriormente manifestou preocupação de que "o artigo não representava adequadamente os dados sobre segurança disponíveis quando o artigo estava sendo revisado para publicação" (17). Além de omitir os dados desfavoráveis, a publicação estabeleceu uma data-limite para relato de eventos adversos cardiovasculares que era anterior à data limite para o relato de eventos adversos gastrintestinais, sem revelar isso à revista ou aos autores acadêmicos do estudo, o que enviesou os resultados em favor do rofecoxibe.

Posteriormente, outro ensaio clínico mostrou que o rofecoxibe causava um número significativamente maior de infartos e acidentes vasculares encefálicos (AVEs) do que o naproxeno (18), e o fabricante voluntariamente retirou o medicamento do mercado.

Em outras publicações influentes, pesquisadores fabricaram ou alteraram intencionalmente os dados, por exemplo, ao mostrarem uma falsa associação entre vacina para sarampo-caxumba-rubéola e autismo em crianças e ao afirmarem ter derivado uma linhagem de células-tronco humanas usando transplante nuclear de células somáticas (19, 20). Situações de má-prática como essas alimentam a desconfiança do público e dos médicos e ameaçam o financiamento público às pesquisas.

Má-conduta científica

O Office for Research Integrity (Departamento de Integridade na

Pesquisa) do governo federal norte-americano define **má-conduta em pesquisa** como fabricação, falsificação e plágio (21).

- **Fabricação** é elaborar, registrar e publicar resultados falsos.
- **Falsificação** é manipular material, equipamento ou procedimentos de pesquisa ou alterar/omitir dados ou resultados de maneira a deturpar os verdadeiros achados do estudo.
- **Plágio** é se apropriar de ideias, resultados ou palavras de outras pessoas omitindo o devido crédito.

Nessa definição federal norte-americana, a má-conduta pressupõe ação intencional; isto é, o pesquisador está ciente de que a conduta é errada. No Caso 14.3, não foi possível comprovar a falsificação intencional dos achados. A má-conduta em pesquisa, por essa definição, exclui enganos e diferenças legítimas de opinião que fazem parte do processo normal de pesquisa. Ela também não aborda outras atitudes impróprias, como publicação dupla, recusa em compartilhar materiais de pesquisa e assédio sexual; as instituições e pesquisa devem lidar com esses problemas com base em outras regulamentações.

Havendo alegação de má-conduta em pesquisa, tanto a agência de fomento quanto a instituição do pesquisador têm a responsabilidade de conduzir, em tempo hábil, inquérito ou investigação apropriados (22). Durante a investigação, tanto os denunciantes quanto os cientistas acusados têm direitos que devem ser respeitados. Os denunciantes devem ser protegidos contra retaliações, e os acusados devem ser informados sobre as acusações e ter oportunidade de responder a elas. As penalidades para má-conduta comprovada em pesquisa incluem suspensão do auxílio de pesquisa e impossibilidade de novas solicitações de auxílio, além de outras sanções administrativas, acadêmicas, criminais ou cíveis.

CASO 14.3 Efeitos adversos cardíacos do rofecoxibe *(continuação)*

Muitos pacientes que haviam tomado rofecoxibe e sofrido infarto processaram o fabricante. Durante o processo legal, houve apreensão de comunicações internas por *e-mail* do financiador do estudo, que

mostraram que muitos artigos sobre o rofecoxibe foram escritos por funcionários ou consultores da indústria farmacêutica, e os pesquisadores acadêmicos foram convidados a serem primeiros autores apenas após a primeira versão do manuscrito já ter sido elaborada. Muitas vezes os funcionários que haviam escrito a primeira versão do manuscrito não foram listados ou reconhecidos como autores na versão final do artigo.

Autoria

Para justificar a autoria, os pesquisadores devem ter feito contribuições substanciais para:

- Concepção e delineamento do estudo, ou análise e interpretação de dados,
- Elaboração ou revisão do manuscrito; e
- Aprovação final do artigo. (23)

A **autoria honorária e a autoria fantasma** não são eticamente aceitáveis. A autoria honorária refere-se àqueles autores que fazem mínimas contribuições ao artigo, como fornecer seu prestígio, acesso a participantes, reagentes, assistência laboratorial ou financiamento. No Caso 14.3, não é adequado tornar-se autor após o estudo ter sido concluído, os dados analisados e a primeira versão do manuscrito ter sido elaborada. A autoria fantasma refere-se àqueles indivíduos que fazem contribuições substanciais ao artigo, mas não são listados como autores. Eles em geral são funcionários de indústrias farmacêuticas ou de empresas de redação científica. Quando os autores fantasmas não são listados, os leitores são levados a subestimar o papel da indústria farmacêutica no manuscrito. De acordo com um estudo, 25% dos artigos em revistas de alto impacto tinham autores honorários e 12% tinham autores fantasmas (24).

Frequentemente surge discordância sobre quem deveria ser listado como autor ou sobre a ordem dos autores. Essas questões devem ser discutidas e decididas antes do início do projeto. Mudanças na autoria deverão ser negociadas se forem decididas alterações nas

responsabilidades pelo trabalho. Sugestões de como proceder diplomaticamente em tais negociações estão disponíveis (25). Uma vez que não há concordância sobre critérios para a ordem dos autores, algumas revistas solicitam que sejam descritas no artigo as contribuições de cada autor.

Conflitos de interesse

Os principais interesses de um pesquisador deveriam ser fornecer respostas válidas a questões científicas importantes e proteger a segurança dos participantes. Os pesquisadores podem também ter outros interesses, como sua reputação ou renda, que podem estar em conflito com os objetivos principais da pesquisa e podem comprometer sua objetividade e alimentar a desconfiança do público em relação à pesquisa (33, 34). O menor indício de um conflito de interesse, mesmo que não fundamentado, pode ser deletério (26).

Tipos de conflito de interesse

- **Conflitos de interesses financeiros.** Estudos com novos medicamentos, dispositivos e testes são normalmente financiados pela indústria. A preocupação ética surge quando os laços financeiros podem influenciar no delineamento e na condução do estudo, na interpretação exagerada de um resultado positivo e na omissão da publicação de resultados negativos (27, 28). Se os investigadores possuírem patentes sobre a intervenção em estudo ou opções de ações da empresa fabricante do medicamento ou dispositivo em estudo, poderão obter lucros consideráveis se o tratamento se revelar eficaz, além da compensação financeira por terem conduzido a pesquisa. Por fim, o pagamento de altos valores em consultorias, honorários ou prêmios em espécie podem enviesar o julgamento de um pesquisador em favor do produto da empresa.

- **Duplo papel médico-investigador.** Se o investigador for o próprio médico de um potencial participante da pesquisa, o papel de clínico e investigador podem entrar em conflito. Os pacientes podem temer que seu atendimento futuro fique comprometido caso se recusem a

participar do estudo e também podem não perceber a diferença entre o atendimento médico e a participação na pesquisa. E o que é melhor para um determinado paciente pode não ser o melhor para um projeto de pesquisa.

Lidando com interesses conflitantes

Todos os conflitos de interesse devem ser revelados e, para alguns, o potencial de enviesar os resultados da pesquisa é tão grande que eles deveriam ser manejados ou evitados.

- **Reduzir a possibilidade de viés.** Um ensaio clínico bem delineado inclui inúmeras precauções-padrão que visam a controlar interesses conflitantes. O investigador pode ser **cego** a que intervenção o participante recebe, evitando, assim, que faça uma avaliação tendenciosa dos resultados. Um **comitê** independente **de monitoramento de dados e da segurança** dos participantes (ver Capítulo 11), cujos membros não tenham conflitos de interesse, pode revisar dados durante o andamento da pesquisa e decidir por abortá-la caso os benefícios ou malefícios sejam convincentemente evidentes. O processo de **revisão por pares** para a concessão de verba de pesquisa e para a publicação de resumos e artigos também contribui para reduzir o viés.

- **Separar os papeis conflitantes.** Cabe ao médico separar o papel de investigador de um projeto de pesquisa daquele de clínico que presta atendimento médico ao participante. Em geral, os médicos não devem incluir seus próprios pacientes em um estudo do qual são coinvestigadores. Se tais pacientes forem incluídos, um outro membro da equipe de pesquisa sem responsabilidade clínica pelo paciente pode encarregar-se das discussões sobre o consentimento.

- **Controlar a análise e as publicações.** Em pesquisas financiadas pela indústria farmacêutica, os investigadores de instituições acadêmicas devem assegurar-se de que o contrato concede **acesso irrestrito aos dados primários e às análises estatísticas**, bem como **liberdade de publicação dos achados**, mesmo nos casos em que o medicamento não

seja eficaz (27, 28). O investigador tem obrigação ética de se responsabilizar por todos os aspectos da pesquisa. O patrocinador tem o direito de revisar os manuscritos, emitir sugestões e assegurar que as inscrições de patentes tenham sido encaminhadas ou preenchidas antes que o artigo seja submetido à publicação. Entretanto, o patrocinador não deverá exercer nenhum tipo de veto ou censura e tampouco insistir em aspectos específicos da redação do manuscrito.

- **Explicitar conflitos de interesse.** As instituições de pesquisa exigem que os conflitos de interesse sejam explicitados a um órgão competente. O NIH e outras agências de financiamento, os CEPs locais, os congressos científicos e as revistas médicas exigem que os autores explicitem os conflitos de interesse ao submeterem propostas de auxílio, resumos ou artigos. Embora isso, por si só, seja uma resposta insuficiente a conflitos de interesses graves, pode coibir práticas difíceis de serem justificadas eticamente e permitir aos revisores e leitores dos artigos avaliar o potencial para influência indevida.

- **Gerenciar os conflitos de interesse.** Se um determinado estudo apresentar importantes conflitos de interesse, a instituição de pesquisa, a agência de financiamento ou o CEP podem exigir certas salvaguardas adicionais, como monitoramento mais próximo do processo de consentimento informado ou modificação do papel do investigador que apresenta o conflito.

- **Proibir certas situações.** Para minimizar os conflitos de interesse, agências de financiamento e instituições acadêmicas podem proibir que detentores de patentes sobre uma intervenção ou pessoas ligadas à empresa que fabrica a intervenção sejam os investigadores principais em um ensaio clínico.

QUESTÕES ÉTICAS ESPECÍFICAS A CERTOS TIPOS DE ESTUDOS

Ensaios clínicos randomizados

Embora o ensaio clínico randomizado seja a forma mais rigorosa de se avaliar as intervenções (Capítulo 10), ele gera preocupações éticas especiais por dois motivos: a alocação do tratamento é determinada pelo acaso e, ao contrário dos estudos observacionais, os pesquisadores realizam uma intervenção nos participantes. Uma justificativa ética para designar aleatoriamente um tratamento é que as intervenções do estudo estão em **equipolência**, um conceito que parece claro em termos intuitivos, porém é muito debatido e impossível de definir com precisão (29). Deve haver incerteza ou controvérsia genuínas sobre qual braço do estudo é superior, de modo que os participantes não serão muito prejudicados se permitirem que seu cuidado seja determinado pela randomização, e não pelo seu médico pessoal. A equipolência não requer um equilíbrio exato entre os braços do estudo.

As pessoas que participam de um ensaio clínico recebem uma intervenção cujos efeitos adversos muitas vezes são desconhecidos. Portanto, deve ser feito um monitoramento cuidadoso para assegurar que os participantes não sofram danos inapropriadamente. É responsabilidade do investigador instituir métodos cuidadosos para avaliar os efeitos adversos (ver Capítulos 10 e 11). Para a maioria dos ensaios clínicos, isso inclui constituir um Comitê de Monitoramento dos Dados e da Segurança (Data and Safety Monitoring Board, DSMB) que revisa de forma intermitente os dados do estudo e tem poder para interrompê-lo caso haja dano inesperado associado à intervenção (ver Capítulo 11).

A escolha da intervenção apropriada para o grupo-controle também gera inquietações éticas. Caso haja um cuidado eficaz que seja considerado padrão para a doença, o grupo-controle deve recebê-lo. Entretanto, controles-placebo podem ser aceitáveis em ensaios clínicos de curta duração que não ofereçam riscos sérios aos participantes, como por exemplo, em estudos de hipertensão leve e dores leves, autolimitadas. Os participantes devem ser informados sobre as intervenções eficazes disponíveis fora do estudo.

É considerado antiético dar continuidade a um ensaio clínico após ter sido demonstrado que uma das terapias em estudo é mais segura ou mais eficaz do que a outra. Além disso, seria errôneo prosseguir com um ensaio clínico que não pudesse responder à questão de pesquisa em prazo aceitável devido ao baixo recrutamento de participantes, à baixa

incidência de desfechos ou à alta taxa de abandonos. Análises periódicas de dados interinos por um comitê independente de monitoramento da segurança e dos dados podem determinar se um ensaio deveria ser suspenso precocemente devido a essas razões (30). Essas análises interinas não devem ser realizadas pelos investigadores do estudo, pois o descegamento para os achados interinos podem resultar em viés se o estudo continuar, e os investigadores muitas vezes têm um conflito de interesse em relação a continuar ou interromper um estudo. Procedimentos para análise de dados de estudos em andamento e regras estatísticas para interrupção dos estudos devem ser especificados antes de os participantes começarem a ser incluídos (ver Capítulo 11).

Ensaios clínicos conduzidos em países em desenvolvimento apresentam dilemas éticos adicionais (Capítulo 18).

Pesquisa com dados e materiais biológicos previamente coletados[12]

As pesquisas com dados ou materiais biológicos previamente coletados oferecem grande potencial para descobertas significativas. Por exemplo, testes de DNA em grandes bancos de soro armazenado “linkados” a dados clínicos poderiam identificar genes que aumentam a probabilidade de desenvolver uma doença, ter um mau prognóstico ou responder a um determinado tratamento. Grandes biotecas de amostras de sangue e de outros materiais biológicos armazenados permitem que futuramente sejam realizados novos estudos sem a necessidade de coletar novas amostras. Esses estudos adicionais não oferecem nenhum risco físico aos participantes. No entanto, podem trazer preocupações éticas. O consentimento para estudos futuros não especificados é complicado, porque não é possível prever o tipo de pesquisa que será realizado posteriormente.

Além disso, os participantes poderiam ter objeções a certos usos futuros dessas amostras e dados. Se ocorrerem quebras de confidencialidade, elas poderão levar a estigma e discriminação. Os grupos que participam de uma pesquisa podem estar sujeitos a danos até mesmo se não houver dano aos participantes individuais.

Quando materiais biológicos forem coletados para uso futuro, os formulários de consentimento deverão permitir que os participantes da

pesquisa concordem ou não com certas categorias gerais para o uso posterior de suas amostras. Como exemplo, os participantes poderão concordar com o uso de suas amostras em pesquisas futuras:

- Para as pesquisas futuras aprovadas por um CEP e um painel de revisão científica;
- Apenas para pesquisas sobre condições específicas; ou
- Apenas para o estudo atual e não para estudos futuros.

Os participantes também devem ser informados sobre se os dados e amostras identificáveis serão compartilhados com outros pesquisadores. Além disso, os participantes devem estar cientes de que descobertas científicas de estudos com amostras da bioteca do estudo poderão ser patenteadas e transformadas em produtos comerciais.

OUTRAS QUESTÕES

Pagamento aos participantes da pesquisa[13]

Os participantes de pesquisas clínicas merecem compensações financeiras por seu tempo, esforço e gastos associados: transporte, pagamento para alguém cuidar de seus filhos, etc. Falando de forma prática, as compensações podem também ser um atrativo para incluir e manter os participantes na pesquisa. É comum oferecer somas mais altas aos participantes de pesquisas de maior inconveniência ou risco. Entretanto, os pagamentos também geram preocupações éticas sobre a indução inapropriada dos participantes a tomarem parte na pesquisa. Se eles receberem mais por participarem de pesquisas arriscadas, pessoas de nível socioeconômico mais baixo poderão ser seduzidas a correrem riscos que julgariam inaceitáveis se estivessem em melhores condições de julgamento. A fim de evitar a indução inapropriada, foi sugerido que as compensações sejam calculadas de acordo com os gastos reais dos participantes e que o tempo dispensado seja pago de acordo com uma tarifa por hora de trabalho não especializado (31).

RESUMO

1. Os investigadores devem garantir que seus projetos observem os **princípios éticos** do **respeito à pessoa,** da **beneficência** e da **justiça.**
2. Os investigadores devem assegurar que as pesquisas estejam em concordância com as **regulamentações federais**. Os aspectos mais importantes são o **consentimento informado** e a **apreciação** pelo **CEP**. Durante o processo de consentimento informado, os investigadores devem explicar aos potenciais participantes a **natureza do projeto** e seus **procedimentos**, o potencial de **riscos**, os **benefícios** e as **alternativas**. Os investigadores devem assegurar a **confidencialidade** das informações dos participantes, observando o **HIPAA** Health Privacy Rule.
3. **Indivíduos vulneráveis**, especialmente **crianças, presidiários, gestantes** e pessoas com **capacidade mental reduzida** ou **desvantagem social,** requerem proteções adicionais.
4. Os investigadores devem ter **integridade ética.** Não podem cometer atos de má-conduta científica, definidos como **fabricação**, **falsificação** ou **plágio**. Devem explicitar e gerenciar adequadamente os **conflitos de interesse** e seguir critérios apropriados para **autoria**, sendo listados como autores em um manuscrito apenas se tiverem feito contribuições intelectuais substanciais, e assegurando que todas as pessoas que contribuíram de forma substancial para o manuscrito sejam listadas como autores.
5. Em certos tipos de pesquisa, outras questões éticas devem ser abordadas. Nos ensaios clínicos randomizados, os braços da intervenção devem estar em **equipolência,** os controles devem receber **intervenções apropriadas** e o estudo não deve ser continuado quando se demonstrou que um dos braços é mais eficaz ou danoso. Quando a pesquisa utilizar materiais biológicos ou dados coletados previamente, é importante dar especial atenção à **confidencialidade**.

REFERÊNCIAS

1. Jones JH. The Tuskegee syphilis experiment. In: Emanuel EJ, Grady C, Crouch RA, et al., editors. *Oxford textbook of research ethics*. New York: Oxford University Press, 2008, 86–96.

2. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of biomedical and behavioral research*. 1979. Available at: www.hhs.gov/ohrp/humansubjects/guidance/belmont.html, accessed 8/27/12.
3. Department of Health and Human Services. *Protection of human subjects 45 CFR part 46*. 2005. Available at: www.dhhs.gov/ohrp/humansubjects/guidance/45cfr46.html, accessed 9/27/12.
4. Emanuel EJ, Menikoff J. Reforming the regulations governing research with human subjects. *N Engl J Med* 2011; 365:1145–50.
5. Lo B, Barnes M. Protecting research participants while reducing regulatory burdens. *JAMA* 2011;306:2260–2261.
6. Department of Health and Human Services. *Protocol review*. 2005. Available at: www.dhhs.gov/ohrp/policy/protocol/index.html, accessed 9/27/12.
7. King NMP, Churchill LR. Assessing and comparing potential benefits and risks of harm. In: Emanuel EJ, Grady C, Crouch RA, et al., editors. *The Oxford textbook of clinical research ethics*. New York: Oxford University Press, 2008, 514–526.
8. Henderson GE, Churchill LR, Davis AM, et al. Clinical trials and medical care: defining the therapeutic misconception. *PLoS Med* 2007;4:e324.
9. Federman DD, Hanna KE, Rodriguez LL. *Responsible research: a systems approach to protecting research participants*. 2002. Available at: www.nap.edu/catalog.php?record_id=10508, accessed 9/29/12.
10. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. *JAMA* 2004;292:1593–1601.
11. Lomax GP, Hall ZW, Lo β. Responsible oversight of human stem cell research: the California Institute for Regenerative Medicine's medical and ethical standards. *PLoS Med* 2007;4:e114.
12. Woodsong C, Karim QA. A model designed to enhance informed consent: experiences from the HIV prevention trials network. *Am J Public Health* 2005;95:412–419.
13. Wolf LE, Dame LA, Patel MJ, et al. Certificates of confidentiality: legal counsels' experiences with perspectives on legal demands for reseasch data. *J Empir Res Hum Res Ethics* 2012;7:1–9.
14. Nass SJ, Leavitt LA, Gostin LO. *Beyond the HIPAA Privacy Rule: enhancing privacy, improving health through research*. 2009. Available at: http://iom.edu/Reports/2009/Beyond-the-HIPAA-Privacy-Rule-Enhancing-Privacy-Improving-Health-Through-Research.aspx, accessed 9/29/12.
15. National Bioethics Advisory Commission. *Ethical and policy issues in research involving human participants*. Rockville, MD: National Bioethics Advisory Commission, 2001.
16. Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal

toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N Engl J Med* 2000;343:1520–1528.

17. Curfman GD, Morrissey S, Drazen JM. Expression of concern. *N Engl J Med* 2005;353:2813–2814.
18. Bresalier RS, Sandler RS, Quan H, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med* 2005;352:1092–1102.
19. Godlee F, Smith J, Marcovitch H. Wakefield's article linking MMR vaccine and autism was fraudulent. *BMJ* 2011;342:c7452.
20. Kennedy D. Responding to fraud. *Science* 2006;314:1353.
21. Office of Research Integrity. *Case summaries*. Available at: http://ori.hhs.gov/case_summary, accessed 9/29/12.
22. Mello MM, Brennan TA. Due process in investigations of research misconduct. *N Engl J Med* 2003;349:1280–1286.
23. International Committee of Medical Journal Editors. *Uniform requirements for manuscripts submitted to biomedical journals*. Available at: www.icmje.org/faq_urm.html, accessed 9/29/12.
24. Wislar JS, Flanagin A, Fontanarosa PB, Deangelis CD. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ* 2011;343:d6128.
25. Browner WS. Authorship. In: *Publishing and presenting clinical research*, 2nd ed. Philadelphia: Lippincott Williams & Willkins, 2006, 137–144.
26. Lo B, Field M. *Conflict of interest in medical research, education, and practice*. 2009. Available at: www.iom.edu/Reports/2009/Conflict-of-Interest-in-Medical-Research-Education-and-Practice.aspx, accessed 11/16/11.
27. DeAngelis CD, Fontanarosa PB. Ensuring integrity in industry-sponsored research: primum non nocere, revisited. *JAMA* 2010;303:1196–1198.
28. DeAngelis CD, Fontanarosa PB. Impugning the integrity of medical science: the adverse effects of industry influence. *Jama* 2008;299:1833–1835.
29. Joffe S, Miller FG. Equipoise: asking the right questions for clinical trial design. *Nat Rev Clin Oncol* 2012;9:230–235.
30. Ellenberg SS, Fleming TR, DeMets DL. *Data monitoring committees in clinical trials*. Chichester, England: Wiley, 2003.
31. Grady C. Payment of clinical research subjects. *J Clin Invest* 2005;115:1681–1687.

---

1 N de R.T. Embora a abordagem das questões éticas deste capítulo tenha sido a partir da problemática de pesquisa norte-americana, ela vale bem para nossa realidade. As diferenças em relação à realidade brasileira são apontadas por meio de notas de rodapé. Mais informações podem ser buscadas na página da Comissão Nacional de Ética em Pesquisa (*http://conselho.saude.gov.br/web_comissoes/conep/index.html*) ou do Núcleo Institucional de Bioética HCPA/UFRGS (*http://www.ufrgs.br/HCPA/gppg/bioetica.htm*).

2 N. de R.T. No Brasil, há menos situações em que há isenção de avaliação pelo Comitê de Ética em Pesquisa. Em caso de dúvida, deve-se questionar o Comitê de Ética em Pesquisa da instituição.

3 N. de R.T. As diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos no Brasil estão descritas na Resolução 466/2012, do Conselho Nacional de Saúde (CNS), disponível em http://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf.

4 N. de R.T. No Brasil, em algumas situações, o CEP deve encaminhar o protocolo para apreciação pela Comissão Nacional de Ética em Pesquisa (CONEP), vinculada ao Conselho Nacional de Saúde (CNS). Alguns exemplos de situações desse tipo incluem: pesquisa genética, reprodução assistida, estudos com financiamento de entidades estrangeiras, estudos em que é feito armazenamento de materiais biológicos e estudos que, a critério do CEP, sejam considerados merecedores de avaliação pela CONEP.

5 N. de R.T. No Brasil, desde 2011, a submissão de protocolos ao CEP é feita por meio da Plataforma Brasil, que pode ser encontrada em http://aplicacao.saude.gov.br/plataformabrasil/. Ela permite que as pesquisas sejam acompanhadas em seus diferentes estágios – desde sua submissão até a aprovação final pelo CEP e pela Conep, quando necessário. Por meio dessa plataforma, podem ser enviados, também, os relatórios parciais e finais das pesquisas.

6 N. de R.T. Não há revisão expedita no sistema brasileiro. Para todos os protocolos de pesquisa, o CEP tem o prazo de até 30 dias para fornecer seu parecer. Quando submetidos ao CONEP, o prazo é de até 60 dias.

7 N. de R.T. Os CEPs podem variar em relação ao formato recomendado para o consentimento informado. Recomenda-se consultar o CEP em relação ao grau de detalhamento necessário.

8 N. de R.T. Um exemplo pode ser o do Hospital de Clínicas de Porto Alegre (HCPA) que tem suas recomendações para redação do termo de consentimento informado em http://www.ufrgs.br/bioetica/TCLEHCPA2011.htm.

9 N. de R.T. A única exceção prevista nas normas brasileiras está contida na Resolução 196/96: "casos em que seja impossível registrar o consentimento livre e esclarecido; tal fato deve ser devidamente documentado com explicação das causas da impossibilidade e parecer do Comitê de Ética em Pesquisa".

10 N. de R.T. Não existe recurso semelhante no Brasil.

11 N. de R.T. No Brasil, a autorização para uso de dados do prontuário eletrônico para fins de pesquisa é feita unicamente por meio de aprovação pelo CEP e da obtenção do Termo de Consentimento Informado Livre e Esclarecido.

12 N. de R.T. No Brasil, a Resolução CNS nº 441, de 12 de maio de 2011, que trata do armazenamento e utilização de material biológico humano com finalidade de pesquisa pode ser encontrada em *http://conselho.saude.gov.br/resolucoes/2011/Reso441.pdf*

13 N. de R.T. No Brasil, a Resolução 466/2012 estipula que a participação na pesquisa deve ser

gratuita, porém abre a possibilidade de pagamento aos participantes nas pesquisas de Fase I ou de bioequivalência, possibilidade essa não prevista na resolução anterior (196/1996). Estipula também que os participantes devem ser ressarcidos pelos custos que tiverem com sua participação.

CAPÍTULO 15

Elaborando questionários, entrevistas e instrumentos *on-line*

Steven R. Cummings, Michael A. Kohn e Stephen B. Hulley

Grande parte das informações utilizadas na pesquisa clínica é coletada por meio de **questionários** em papel ou em formato eletrônico, ou por meio de **entrevistas**. A validade dos resultados depende da qualidade desses **instrumentos**. Neste capítulo, descreveremos os componentes de bons questionários e entrevistas e abordaremos os procedimentos para sua elaboração.

Cada vez mais os pesquisadores clínicos têm acesso a opções para desenvolver **questionários *on-line***, como o REDCap, uma plataforma *on-line* de gerenciamento de dados desenvolvida por um consórcio liderado pela Universidade de Vanderbilt, além de produtos comerciais, incluindo o SurveyMonkey, o Zoomerang, o Qualtrics e o QuesGen. Esses produtos oferecem ferramentas e funcionalidades *on-line* para facilitar o desenvolvimento de instrumentos de coleta de dados que permitem o envio de *e-mails* automáticos para os participantes e a postagem na página do estudo na internet. Apesar dessa transição, ainda em curso, de questionários em papel para instrumentos *on-line*, os princípios da elaboração de bons instrumentos permanecem os mesmos: escrever instruções claras e questões bem formuladas que estimulam respostas informativas (1).

ELABORANDO BONS INSTRUMENTOS

Questões abertas e questões fechadas

Existem dois tipos básicos de questões, com diferentes propósitos: questões abertas e questões fechadas. As **questões abertas** são utilizadas quando o objetivo é permitir que o respondente utilize suas próprias palavras para respondê-las. A pergunta a seguir é um exemplo de questão aberta:

> Quais são os hábitos que, na sua opinião, aumentam a probabilidade de um indivíduo ter um derrame?

As questões abertas dão mais liberdade ao respondente, com menos limites impostos pelo investigador. Elas permitem que os participantes forneçam mais informações do que seria possível a partir de uma lista fechada de respostas, porém as respostas podem ser menos completas. A grande **desvantagem** das questões abertas é que elas requerem métodos qualitativos ou sistemas especiais para codificação e análise das respostas (p. ex., dicionários para codificação dos sintomas e eventos adversos); isso consome mais tempo do que registrar as respostas a questões fechadas e pode exigir maior julgamento subjetivo. Esse tipo de questão costuma ser usado na fase exploratória do processo de elaboração das questões, pois ajuda o pesquisador a compreender os conceitos na forma expressa pelos respondentes. As frases e palavras por eles usadas podem formar a base para as **questões fechadas** que pedem que o respondedor selecione a partir de duas ou mais respostas pré-selecionadas:

> Quais das alternativas a seguir, na sua opinião, aumentam mais a probabilidade de um indivíduo ter um derrame? (Assinale todas as alternativas corretas.)
> ☐ Fumo
> ☐ Excesso de peso
> ☐ Estresse
> ☐ Consumo de álcool

Uma vez que as **questões fechadas** oferecem uma lista de respostas possíveis, elas são **mais rápidas e fáceis de responder**, e as respostas são **mais fáceis de tabular** e analisar. Além disso, a lista de possíveis respostas geralmente facilita o entendimento das questões, e as questões

fechadas podem ter como alternativas de resposta escalas de itens múltiplos que produzem um único escore.

Por outro lado, as questões fechadas têm inúmeras **desvantagens**. Elas conduzem os respondentes a certas direções, não permitindo que eles formulem sua própria resposta, potencialmente mais correta. As respostas possíveis listadas pelo investigador podem não ser **exaustivas** (não incluir todas as opções possíveis, p. ex., "atividade sexual" ou "sal na dieta"). Uma solução é incluir uma opção como *Outra(s) (favor especificar)* ou *Nenhuma das anteriores*. Quando o objetivo for o de se obter uma única resposta, o respondente deverá ser informado sobre isso, e o conjunto de alternativas possíveis terá de ser **mutuamente exclusivo** (i.e., as categorias não deverão se sobrepor), garantindo, assim, clareza e parcimônia.[1]

Quando a questão der margem a mais de uma resposta, não é aconselhável instruir que o respondente marque "todas as que se aplicam", pois esse procedimento não exige que ele considere cada uma das possíveis alternativas. Um item não considerado pode ser uma resposta que não se aplica ou um item que passou despercebido. Para evitar esse problema, o melhor a fazer é pedir aos respondentes que marquem "sim" ou "não" para cada alternativa possível:

Quais das alternativas a seguir aumentam a probabilidade de um indivíduo ter um derrame?

	Sim	**Não**	**Não sei**
Fumo	○	○	○
Excesso de peso	○	○	○
Estresse	○	○	○
Consumo de álcool	○	○	○

As respostas para questões fechadas também podem ser registradas usando linhas ou outras representações gráficas em uma **escala visual analógica (EVA).** O participante é solicitado a marcar, ao longo de uma linha contínua, o ponto que melhor representar sua resposta. É importante que as palavras-chave em cada extremidade descrevam os valores extremos do item de interesse. A seguir, apresentamos uma EVA para intensidade da dor:

Por favor, marque na linha abaixo o ponto que melhor descreve a intensidade de sua dor ao longo da última semana.

Nenhuma ____ **Insuportável**

Por conveniência, as linhas costumam medir 10 cm, e o escore será a distância em centímetros da menor extremidade. Para um exemplo de uma EVA *on-line*, veja o *site* http://www.epibiostat.ucsf.edu/dcr.

As escalas visuais analógicas são uma opção atraente, pois medem características em uma escala contínua; elas são mais sensíveis a mudanças pequenas do que as medições baseadas em listas de adjetivos categóricos. Muitas das ferramentas para construção de questionários *on-line*, incluindo o REDCap, o Qualtrics e o QuesGen, permitem o uso de EVAs.

Formatação

Nos questionários, geralmente é apresentada no início uma breve descrição dos objetivos do estudo e de que maneira os dados serão usados. Informação similar é geralmente apresentada no início de uma entrevista como parte do processo de obtenção de consentimento. A fim de assegurar respostas acuradas e padronizadas, todos os instrumentos devem trazer instruções especificando como deve ocorrer o seu preenchimento. Esse procedimento vale não somente para questionários de autopreenchimento, mas também para os formulários usados pelos entrevistadores para o registro das respostas.

Às vezes é útil fornecer um exemplo de como responder a uma questão, usando uma questão simples, de fácil resposta.

Instruções para o Preenchimento de um Questionário para Avaliação da Ingestão Alimentar

Estas questões abordam seus hábitos alimentares ao longo dos últimos 12 meses. Por favor, marque, ao lado do nome dos alimentos, a quantidade habitual e com que frequência você ingere cada um deles.

Por exemplo, se você toma um copo médio (180 mL) de suco de maçã três vezes por semana, você responderia:

Suco de maçã ○ Pequeno (90 mL) [3] vez(es) por ○ Dia

⦿ Médio (180 mL) ○ Grande (270 mL)	⦿ Semana ○ Mês ○ Ano

Para melhorar o fluxo do instrumento, podem-se agrupar as questões em grupos temáticos, introduzindo-as com cabeçalhos ou breves descrições. Para preparar o respondente, é aconselhável iniciar a entrevista com perguntas emocionalmente neutras, por exemplo sobre seu nome e informações para contato. Perguntas muito delicadas, como sobre renda ou função sexual, são geralmente deixadas para o final do instrumento. Para cada questão ou conjunto de questões com formato diferente das outras questões, novas instruções devem indicar claramente a maneira de respondê-las.

Se as questões incluírem intervalos de tempo diferentes, é aconselhável situar o paciente no intervalo desejado, posicionando o intervalo de tempo no topo de cada conjunto de perguntas. Por exemplo, questões tipo:

Quantas vezes você consultou um médico durante o último ano?
Quantas vezes você foi atendido em um serviço de emergência durante o último ano?
Durante o último ano, quantas vezes você foi admitido em algum hospital?

podem ser simplificadas da seguinte maneira:

Durante o último ano, quantas vezes você

- Consultou um médico?
- Foi atendido em um serviço de emergência?
- Foi admitido em um hospital?

Para **formulários em papel**, a **diagramação** deve facilitar o máximo possível que os respondentes, sejam eles os participantes do estudo ou integrantes da equipe de pesquisa, completem todas as questões na sequência correta. Se o formato for muito complexo, os respondentes ou os entrevistadores poderão pular questões, fornecer dados errados e, até mesmo, recusar-se a completar o instrumento. Um questionário **organizado** e com **espaço suficiente** para as respostas é mais atraente e

mais fácil de usar do que um questionário abarrotado e desordenado. Embora um questionário possa parecer mais curto se tiver menos páginas, sua aplicação poderá ser bem mais difícil quando muitas questões se amontoarem em uma única página. Respostas em escalas devem ser espaçadas adequadamente, para que seja fácil circular ou marcar o número correto sem incluir acidentalmente, também, a resposta acima ou abaixo. Quando forem incluídas questões abertas, o espaço para a resposta deverá ser suficiente para acomodar textos com letras muito grandes. Pessoas com problemas de visão, incluindo idosos, dão preferência a textos com letras grandes e cores contrastantes (como preto e branco).

As respostas possíveis para questões fechadas devem estar alinhadas verticalmente e precedidas por quadrados, círculos ou parênteses para se assinalar ou por números para se circular, em vez de utilizar espaços em branco:

Quantos medicamentos diferentes você toma diariamente?
○ Nenhum
○ 1-2
○ 3-4
○ 5-6
○ 7 ou mais

Observe que essas opções de resposta são exaustivas e mutuamente exclusivas.

Às vezes o investigador pode querer se aprofundar em alguma resposta, elaborando questões mais detalhadas. A melhor maneira de fazer isso é usar uma **questão ramificada**. As respostas dos entrevistados a uma pergunta inicial, geralmente denominada **filtro**, determinam se eles serão direcionados a responder a perguntas adicionais ou a saltar para questões mais adiante. Por exemplo:

Já lhe disseram que você tem pressão alta?
○ Sim → Quantos anos você tinha quando lhe disseram pela primeira vez que você tem pressão alta?
_ _ Anos
○ Não

↓
Vá para a questão 11

As questões ramificadas poupam tempo e evitam que os entrevistados respondam a perguntas irrelevantes e redundantes. Os participantes são direcionados para a próxima questão pertinente por meio de setas e incluindo instruções tipo *Vá para a questão 11* (veja Apêndice 15).

Os **questionários *on-line*** são geralmente mais claros e fáceis de responder, pois incorporam a lógica dos "pulos". Um sujeito do sexo masculino não verá perguntas sobre história gestacional e apenas verá a questão sobre número de maços-ano se tiver respondido "sim" à pergunta sobre se fuma cigarros. (Ver www.epibiostat.ucsf.edu/dcr/.) Entretanto, a lógica dos pulos precisa ser cuidadosamente validada durante a fase de pré-testes do estudo. Lógicas de pulo complexas podem resultar em questões nas quais falta a questão seguinte e em questões órfãs, que nunca são alcançadas. Uma diagramação adequada, que leva em consideração a possível dificuldade de visão de alguns participantes[2], é tão importante para formulários *on-line* quanto para formulários em papel.

Redação

Cada palavra em uma questão pode influenciar a validade e a reprodutibilidade das respostas. O objetivo é formular questões que sejam simples e livres de ambiguidade, encorajando respostas acuradas e honestas, sem constranger ou ofender os participantes.

- **Clareza.** Assegure-se de que as questões sejam tão claras e específicas quanto possível. Devem-se preferir palavras concretas a termos abstratos. Por exemplo, a pergunta "Em geral, quanto de atividade física você pratica?" é menos clara do que "Em uma semana, você realiza quantas horas de caminhadas rápidas?".

- **Simplicidade.** Utilize palavras e estruturas sintáticas simples e comuns que transmitam claramente uma ideia, evitando termos técnicos e jargões. Para a maioria das pessoas, é mais fácil compreender a expressão "remédios que você pode comprar sem receita médica

especial", em vez de "medicamentos não controlados".

- **Neutralidade.** Evite palavras "pesadas" e estereótipos que sugiram uma resposta. Perguntar "No último mês, quantas vezes você bebeu demais?" pode desencorajar os respondentes a admitirem que eles bebem em excesso. "No último mês, quantas vezes você ingeriu mais de cinco doses em um dia?" é uma questão mais factual, desprovida de julgamentos e menos ambígua.

É importante estabelecer uma atmosfera que permita ao respondente admitir comportamentos e atitudes que possam ser considerados indesejáveis. Por exemplo, quando perguntar ao paciente sobre sua adesão às prescrições médicas, o entrevistador (ou o questionário) pode usar uma introdução: "Às vezes, as pessoas se esquecem de tomar os medicamentos que o médico prescreve. Isso acontece de vez em quando com você?". No entanto, esse tipo de introdução pode gerar um efeito inverso. É importante que os respondentes se sintam à vontade para admitir certos comportamentos, mas o questionário não deve encorajá-los ao exagero.

Coletar informações sobre áreas potencialmente **delicadas,** como comportamento sexual ou renda, é particularmente difícil. Algumas pessoas se sentem mais à vontade em responder a esses tipos de perguntas por meio de questionários de autopreenchimento do que em entrevistas. Entretanto, um entrevistador habilidoso pode, às vezes, obter respostas abertas e honestas. É interessante que se deixe as perguntas potencialmente embaraçosas escritas em um cartão. Assim, o entrevistado poderá responder à questão simplesmente apontando para a resposta.

Estabelecendo unidades de tempo

Para que se possa medir a frequência de um comportamento, é essencial que o respondente se baseie em alguma **unidade de tempo**. Se o comportamento for o mesmo dia após dia, como tomar um comprimido de diurético todas as manhãs, a questão pode ser bem simples: "Quantos comprimidos você toma por dia?".

Muitos comportamentos mudam diariamente, de acordo com as estações ou anualmente. Para medi-los, o investigador deve primeiramente decidir qual aspecto do comportamento é mais importante

para o estudo: **a média** ou **os extremos**. Em um estudo sobre o efeito do álcool no risco de doenças cardiovasculares, pode ser necessário medir o consumo médio ao longo do tempo, porém um estudo sobre a influência do álcool na ocorrência de traumas precisará documentar com que frequência o respondente ingeriu uma quantidade de álcool suficiente para ficar intoxicado.

Questões sobre comportamentos usuais podem ser abordadas de duas maneiras: perguntar sobre comportamento “usual ” ou “típico” ou contar instâncias do comportamento durante um certo período. Por exemplo, pode-se determinar o consumo médio de cerveja pedindo-se para os respondentes estimarem a quantidade de bebida que eles ingerem normalmente.

> Aproximadamente quantas cervejas você toma em uma semana típica (uma cerveja é igual a uma latinha, uma garrafa de 350 mL ou um copo grande)?
> [__] cervejas por semana

Esse formato é simples e conciso. Entretanto, ele parte do pressuposto de que o respondente será capaz de expressar seu comportamento habitual em uma simples estimativa. Como os padrões de ingestão de álcool mudam frequente e substancialmente até mesmo ao longo de pequenos intervalos de tempo, o respondente pode ter dificuldade para decidir no que constitui uma semana típica. Quando questionadas sobre seu padrão usual ou típico, as pessoas tendem a revelar o que fazem com mais frequência, ignorando os extremos. Perguntar sobre ingestão de bebidas em dias normais, por exemplo, subestimará o consumo de álcool caso o respondente beba grandes quantidades nos finais de semana.

Uma abordagem alternativa é quantificar o grau de exposição durante um **período especificado de tempo**.

> Nos últimos sete dias, quantas cervejas você tomou (uma cerveja é igual a uma latinha, uma garrafa de aproximadamente 350 mL ou um copo grande)?
> [__] cervejas nos últimos 7 dias

O objetivo aqui é perguntar sobre um período de tempo recente, o mais

curto possível, que represente fielmente a característica em estudo ao longo de todo o período relevante para a questão de pesquisa. O melhor período de tempo depende da característica em estudo. Por exemplo, os padrões de sono podem variar consideravelmente de um dia para outro, mas questões sobre os hábitos de sono durante a última semana podem representar adequadamente os padrões de sono durante todo o ano. Por outro lado, a frequência da prática de sexo sem proteção varia imensamente de semana para semana, de modo que questões sobre esse assunto devem abranger intervalos de tempo bem maiores.

O uso de **diários** pode ser uma abordagem mais acurada para se acompanhar eventos, comportamentos ou sintomas que ocorrem de forma episódica (como *quedas*) ou que variam a cada dia (como sangramento vaginal). Isso pode ser extremamente útil nos casos em que é importante estabelecer o momento exato no tempo em que os eventos ocorrem ou sua duração, ou quando os eventos de interesse são facilmente esquecíveis. Os participantes podem entrar esses dados em dispositivos eletrônicos, e essa abordagem permite calcular um escore médio diário do evento ou do comportamento em avaliação. No entanto, essa abordagem também consome tempo dos participantes e pode levar a uma maior frequência de dados faltantes do que a abordagem comum de fazer perguntas retrospectivas. O uso de diários pressupõe que o período de tempo avaliado tenha sido típico, e que a conscientização induzida pelo preenchimento de diários não tenha alterado de forma importante o comportamento que está sendo registrado.

Evitando dificuldades de resposta

- **Questões que juntam duas perguntas em uma só.** Cada questão deve conter um único conceito. Considere esta questão que foi elaborada para avaliar o consumo de cafeína: “Quantas xícaras de café ou de chá você toma por dia?” Café contém muito mais cafeína do que chá e ambos diferem em outros aspectos. Assim, uma resposta que combina as duas bebidas não será tão precisa quanto poderia ser. Quando uma questão for elaborada com o objetivo de estimar duas coisas ao mesmo tempo, é melhor dividi-la em duas perguntas distintas. “(1) Quantas xícaras de café você toma em um dia normal?” e “(2) Quantas xícaras

de chá você toma em um dia normal?”

- **Pressupostos implícitos.** Algumas questões apresentam pressupostos que não se aplicam a todas as pessoas que participam de um estudo. Por exemplo, um item sobre depressão pergunta com que frequência, na última semana, “me senti como se eu não fosse conseguir sair da fossa mesmo com a ajuda da minha família”. Isso supõe que o respondente tenha familiares e peça apoio emocional; para aqueles que não têm família ou não procuram o apoio de seus familiares, é difícil responder à questão.

- **As questões e suas alternativas de respostas não combinam.** É importante que as perguntas combinem com as respostas, uma tarefa que pode parecer fácil, mas que frequentemente é feita de maneira incorreta. Por exemplo, uma questão como “Você sentiu alguma dor na semana passada?” não deve ter como respostas as alternativas “nunca”, “raramente”, “frequentemente” e “muito frequentemente”. (A questão deveria ser modificada para “Com que frequência você sentiu dores na semana passada?”, ou a resposta deveria ser trocada para “*sim*” ou “*não*”.) Outro problema muito comum ocorre com as questões sobre intensidade que oferecem as alternativas concordo/discordo. Por exemplo, a seguinte afirmativa é dada ao respondente “Às vezes fico deprimido”; ele é solicitado a responder com “concordo” ou “discordo”. Discordar da afirmativa pode significar que o indivíduo está frequentemente deprimido ou então nunca está deprimido. Uma questão simples sobre a frequência em que a pessoa se sente deprimida, combinada com alternativas sobre frequência (*nunca*, *às vezes*, *frequentemente*), originará uma resposta mais clara.

Escalas e escores para medir variáveis abstratas

É difícil avaliar quantitativamente, por meio de uma única questão, um **conceito abstrato**, como qualidade de vida. Assim, características abstratas são geralmente medidas por escores gerados por uma série de questões organizadas em uma escala (2, 3).

O uso de itens múltiplos para avaliar um conceito apresenta vantagens adicionais sobre questões únicas e sobre uma série de questões

apresentadas de maneiras diferentes que não podem ser combinadas. Comparadas às abordagens alternativas, as **escalas de itens múltiplos** podem incrementar a abrangência das possíveis respostas (p. ex., uma escala de itens múltiplos para qualidade de vida pode gerar escores que variam de + a 100, e uma única questão para avaliar a qualidade de vida pode produzir quatro a cinco respostas, variando de “regular” a “excelente”). Uma desvantagem dessas escalas é que elas podem produzir resultados que dificultem o entendimento intuitivo da medida (p. ex., *qualidade de vida = 46,2*).

Escalas tipo Likert são comumente usadas para quantificar atitudes, comportamentos e domínios de qualidade de vida relacionada à saúde. Essas escalas fornecem aos respondentes uma lista de proposições ou questões e lhes pedem que estimem o grau de sua resposta. A cada resposta é atribuído um número de pontos. Por exemplo, considere um questionário para medir o quanto uma pessoa concorda que uma dieta rica em frutas, verduras e legumes está associada a uma saúde melhor:

Para cada item, circule o número que melhor representa sua opinião:

	Concordo fortemente	Concordo	Sou neutro	Discordo	Discordo fortemente
a. Comer mais frutas, verduras e legumes reduz o risco de doença cardíaca.	1	2	3	4	5
b. Os vegetarianos são mais saudáveis do que as pessoas que comem carne.	1	2	3	4	5
c. Aumentar o consumo de frutas, verduras e legumes retarda o envelhecimento.	1	2	3	4	5

O investigador poderá computar o **escore** total das perguntas de um respondente simplesmente somando o escore de cada item ou calculando a média dos pontos para todos os itens respondidos. Por exemplo, uma pessoa que responde concordar plenamente que comer mais frutas, verduras e legumes reduz o risco de doença cardíaca (um ponto) e que vegetarianos são mais saudáveis do que pessoas que comem carne (um ponto), mas discorda que aumentar o consumo de frutas, verduras e legumes retarda o envelhecimento (4 pontos), faria um escore de 6 pontos. A simples soma ou média dos escores supõe que todos os itens têm o mesmo peso e cada item mede a mesma característica geral.

A **consistência interna** de uma escala pode ser testada estatisticamente

usando-se, por exemplo, medidas como a do **alfa de Cronbach** (4), que estima a consistência global de uma escala. O alfa de Cronbach é calculado a partir de correlações entre escores de itens individuais. Valores acima de 0,80 são considerados excelentes, e valores abaixo de 0,50 são considerados inaceitáveis. Valores baixos para consistência interna indicam que alguns dos itens individuais podem estar medindo características diferentes.

Criando novas escalas

Quando um investigador precisa medir uma característica, mas não existem questionários ou abordagens de entrevistas já estabelecidos, pode ser necessário desenvolver um novo instrumento ou escala. A tarefa pode variar desde a elaboração de uma única questão sobre uma variável secundária em um estudo pequeno até o desenvolvimento e teste de novas escalas de itens múltiplos para medição de um desfecho principal em um estudo multicêntrico. Partindo do lado mais simples, para se elaborar uma questão de importância secundária, basta que se use o bom senso e se apliquem princípios básicos para a redação de bons questionários. Desenvolvida a questão, ela pode ser pré-testada, a fim de garantir que seja clara e produza respostas apropriadas. No outro extremo, para se desenvolver um instrumento sobre um conceito importante, é preciso uma abordagem sistemática que pode levar anos desde o esboço inicial até o produto final.

Essa última abordagem costuma iniciar pela geração de potenciais itens para o instrumento a partir de entrevistas individuais e **grupos focais** (grupos pequenos de pessoas relevantes para a questão de pesquisa que são convidados a passar + ou 2 horas discutindo tópicos específicos referentes ao estudo com um líder de grupo). A seguir, é elaborado o esboço do instrumento, sendo o próximo passo uma revisão crítica por pares, mentores ou peritos no assunto. O investigador, então, dá continuidade com a sequência iterativa de pré-testes, revisão, redução do número de questões e validação – procedimentos que serão descritos na próxima seção (e ilustrados no Exemplo 15.1).

EXEMPLO 15.1 Desenvolvimento de um novo instrumento de itens múltiplos

O National Eye Institute Visual Function Questionnaire ilustra a tarefa meticulosa de elaboração e de testagem de um instrumento de itens múltiplos. Mangione e colaboradores dedicaram vários anos à criação e à testagem da escala que tencionava ser o principal instrumento de medição para os desfechos de muitos estudos sobre doenças oculares (5-6). Eles começaram entrevistando pacientes portadores de doenças oftálmicas sobre a maneira como a doença afetava sua vida. A seguir, entrevistaram grupos focais de pacientes com essas doenças e analisaram as transcrições dessas entrevistas para poder escolher questões e alternativas de respostas relevantes. Com essas informações, produziram e pré-testaram um longo questionário que foi então administrado a centenas de pacientes em vários outros estudos. Eles usaram posteriormente os dados desses estudos para identificar os itens que mais contribuíam para a variação dos escores de pessoa para pessoa e para reduzir o questionário de 51 para 25 itens.

Como a criação e validação de um instrumento de itens múltiplos consome muito tempo, isso deve ser feito apenas para variáveis fundamentais para o estudo ou quando as opções existentes são inadequadas ou inapropriadas aos indivíduos incluídos no estudo.

PASSOS NA MONTAGEM DOS INSTRUMENTOS PARA O ESTUDO

Elaborando uma lista de variáveis

Antes de delinear uma entrevista ou um questionário, deve-se escrever uma lista detalhada das informações que serão coletadas e dos conceitos que serão medidos no estudo. Para cada item, pode-se considerar listar seu papel na análise das principais questões da pesquisa, por exemplo, se é preditor, desfecho ou potencial confundidor.

Preferindo instrumentos de medição já existentes, se adequados

É recomendável organizar um arquivo com questões e instrumentos de medição já existentes para cada variável. Quando vários métodos alternativos de medição estão disponíveis, um procedimento bastante

prático é criar uma pasta eletrônica para cada variável a ser medida e, então, localizar e arquivar cópias de questões e instrumentos para cada um dos itens. É importante que, para a medição dos principais preditores e desfechos, sejam usados os melhores instrumentos possíveis. Desse modo, todo o empenho para a coleta de instrumentos alternativos deverá se concentrar nessas **variáveis principais**.

Pode-se começar coletando instrumentos de outros investigadores que já conduziram estudos que incluíram as medidas de interesse. Questionários existentes e informações sobre sua validade, consistência interna e confiabilidade podem ser encontrados na seção de métodos dos artigos publicados e buscando na internet termos-chave como questionários sobre desfechos em saúde (*health outcomes questionnaires*).

Usar instrumentos de outros estudos tem como vantagem a economia de tempo no desenvolvimento e permite a comparação dos resultados entre os diferentes estudos. O ideal seria poder usar instrumentos já existentes sem nenhuma modificação. Entretanto, se certos itens forem inadequados (como pode ocorrer quando um questionário desenvolvido para um determinado grupo cultural é aplicado a um grupo diferente), será necessário excluí-los, modificá-los ou adicionar outros itens.

Se um instrumento já reconhecido cientificamente for muito extenso, pode ser útil entrar em contato com os seus criadores e verificar a existência de uma versão mais compacta. Excluir itens de uma escala já estabelecida põe em risco o significado dos escores, além de impossibilitar a comparação dos resultados com aqueles baseados na escala intacta. Abreviar uma escala pode também diminuir sua reprodutibilidade ou sensibilidade para a detecção de mudanças. Contudo, pode ser possível excluir seções ou subescalas que não sejam essenciais ao estudo e conservar outras partes intactas.

Elaborando um novo instrumento, se necessário

A primeira versão de um instrumento deve ter um escopo mais amplo, incluindo mais questões sobre o tópico em estudo do que o instrumento terá no final. O investigador deve ler a primeira versão cuidadosamente, tentando responder a cada questão como se ele mesmo fosse o respondente e tentando imaginar como elas poderiam ser mal interpretadas. O objetivo aqui é identificar palavras ou frases confusas ou

de difícil interpretação, localizar palavras abstratas ou jargões que possam ser substituídos por termos mais simples e concretos e determinar se questões complexas podem ser divididas em duas ou mais questões. É aconselhável que colaboradores e pessoas com experiência na elaboração de questionários ajudem a revisar o instrumento. Além de considerar o conteúdo dos itens, deve-se avaliar a sua clareza.

Revisando e abreviando os instrumentos do estudo

Os estudos geralmente coletam mais dados do que serão analisados. Entrevistas, questionários e exames longos podem cansar os respondentes e consequentemente reduzir a acurácia e reprodutibilidade das respostas. É geralmente melhor resistir à tentação de, "já que está sendo feita uma pesquisa", incluir questões ou medidas adicionais. As questões que não são essenciais para responder à questão principal de pesquisa aumentam o esforço despendido na obtenção, na entrada, na limpeza e na análise dos dados. O tempo despendido em dados desnecessários ou pouco importantes pode desviar os esforços e reduzir a qualidade geral e a produtividade do estudo.

Para decidir se um conceito é essencial, uma forma prática é imaginar como serão a análise e a publicação dos resultados. Fazer um esboço das tabelas finais assegura que todas as variáveis necessárias estejam incluídas e ajuda a identificar aquelas menos importantes. Após fazer isso, há uma máxima para decidir quais itens incluir: **Na dúvida, não inclua**.

Pré-testando

Deve-se pré-testar o instrumento quanto à sua clareza e duração. Para as medidas principais, pré-testes extensos podem ser de grande utilidade para avaliar se uma questão produz uma faixa adequada de respostas e para testar a validade e reprodutibilidade do instrumento.

Validando

Os questionários e as entrevistas podem ser avaliados quanto à sua validade (um aspecto da acurácia) e à sua reprodutibilidade (precisão) da mesma forma como é feito com qualquer outro tipo de medição (Capítulo 4). O processo tem início com a escolha de questões que apresentem **validade aparente** (***face validity***), uma avaliação subjetiva, porém

importante para saber se os itens estão realmente avaliando as características que o investigador deseja avaliar. Então, são avaliadas a **validade de conteúdo** (***content validity***) e a **validade de construto** (***construct validity***). Sempre que for factível, instrumentos novos podem ser comparados com um **padrão-ouro** das condições de interesse. Finalmente, a **validade preditiva** (***predictive validity***) de um instrumento pode ser avaliada por meio da correlação das medidas com desfechos futuros.

Caso um instrumento tenha como objetivo medir uma mudança, sua capacidade de detectar essa mudança pode ser testada aplicando-se o instrumento antes e depois de os pacientes receberem um tratamento considerado eficaz por outras medidas. Por exemplo, um instrumento novo delineado para medir a qualidade de vida em pessoas com acuidade visual reduzida poderia incluir questões com validade aparente "Você consegue ler jornal sem usar óculos ou lentes de contato?". Essas respostas poderiam ser comparadas com as respostas obtidas por um instrumento válido já existente (Exemplo 15.1) aplicado a pacientes com cataratas severas e em indivíduos com exame de acuidade visual normal. A capacidade de o instrumento detectar mudança de acuidade visual pode ser testada pela comparação das respostas de pacientes portadores de cataratas antes e depois da cirurgia. Fica claro que o processo de validação de novos instrumentos, além de consumir tempo, é oneroso e somente valerá a pena se instrumentos já existentes forem inadequados para a questão de pesquisa ou para a população a ser estudada.

APLICANDO OS INSTRUMENTOS

Questionários *versus* entrevistas

As duas abordagens básicas para coletar dados sobre atitudes, comportamentos, conhecimentos, história clínica e história pessoal são questionários – instrumentos preenchidos pelos próprios respondentes – e entrevistas – que são aplicadas verbalmente por um entrevistador. Cada uma das abordagens tem suas vantagens e desvantagens.

Os **questionários** são geralmente a maneira mais eficiente e uniforme de se administrar questões simples, como idade ou hábito de fumar. Os

questionários são menos onerosos do que as entrevistas, pois requerem menos tempo da equipe de pesquisa, além de serem mais fáceis de padronizar. As **entrevistas** são, em geral, a melhor abordagem para se obter respostas para questões complicadas que requerem explicações ou orientação. Além disso, elas asseguram que os instrumentos sejam respondidos integralmente. As entrevistas podem ser necessárias quando os participantes apresentam habilidades variadas de leitura e de entendimento das questões. Contudo, as entrevistas são mais caras e consomem mais tempo, além de as respostas poderem ser influenciadas pela relação entre o entrevistador e o respondente.

Ambos os instrumentos podem ser padronizados, porém, as entrevistas são inevitavelmente aplicadas de forma pelo menos um pouco diferente a cada vez. Os dois métodos de coleta de informação são suscetíveis a erros causados por recordação imperfeita de dados; ambos são afetados, embora não necessariamente no mesmo grau, pela tendência dos respondentes a emitirem respostas socialmente aceitáveis.

Entrevistas

A habilidade do entrevistador pode ter grande impacto na qualidade das respostas. A **padronização** dos procedimentos de uma entrevista para outra é a chave para maximizar a reprodutibilidade, com uniformidade do palavreado e dos sinais não verbais. Os entrevistadores devem se esforçar para evitar mudanças de palavras ou de tom de voz para não introduzir seus próprios vieses nas respostas. Para que o entrevistador se sinta confortável em ler as questões oralmente, a entrevista deve ser escrita em uma linguagem que lembre a linguagem falada habitual. Questões que soam artificiais ou muito formais encorajam os entrevistadores a improvisar seu próprio estilo de indagar as questões, mais natural, porém menos padronizado.

Às vezes é necessário ajudar um entrevistado a chegar a uma resposta mais apropriada ou a esclarecer o significado de uma resposta. Esse **esclarecimento** também pode ser padronizado simplesmente escrevendo frases-padrão nas margens ou abaixo de cada questão. Por exemplo, uma questão perguntando quantas xícaras de café o respondente ingere em um dia normal pode originar algumas respostas tipo “Eu não sei ao certo, muda de um dia para o outro”. Para casos como esse, o instrumento

poderá conter um esclarecimento-padrão do tipo “Faça o possível para se lembrar; diga aproximadamente quantas xícaras de café você toma em um dia comum”.

As entrevistas podem ser conduzidas pessoalmente ou por telefone. As **entrevistas por telefone assistidas por computador** (*computer assisted telephone interview*, CATI) são uma técnica de inquérito telefônico na qual o entrevistador segue um roteiro e o computador facilita a coleta e edição dos dados. Sistemas de **resposta interativa por voz** (*interactive voice response*, IVR) substituem o entrevistador por questões geradas pelo computador que coletam as respostas dos sujeitos por meio do teclado do telefone ou de reconhecimento de voz (7). Entretanto, **entrevistas presenciais** podem ser necessárias se o estudo exigir observação direta ou exames físicos dos participantes, ou se os possíveis participantes não tiverem telefone (p. ex., moradores de rua).

Métodos de aplicação de questionários

Os questionários podem ser aplicados nos sujeitos pessoalmente, pelo correio, por *e-mail* ou por uma página na internet. A distribuição dos questionários pessoalmente permite que o pesquisador explique as instruções antes que os sujeitos comecem a responder às questões. Quando o estudo requer que o participante se dirija ao local da pesquisa para exames, os questionários podem ser enviados previamente à visita, e todas as respostas devem ser verificadas quanto à sua completude antes que o participante se retire.

Os **questionários enviados por *e-mail*** têm inúmeras vantagens sobre aqueles enviados pelo correio. Embora somente possam ser enviados a participantes com acesso à internet e que tenham familiaridade com ela, os questionários enviados por *e-mail* são uma forma prática de fornecer dados que podem ser incluídos diretamente nas bases de dados.

Questionários publicados em páginas na internet ou disponíveis em **computadores de mão** vêm ganhando espaço como abordagens eficientes e baratas para coletar informações sobre saúde (8). Eles podem produzir dados com erro mínimo, pois é possível verificá-los automaticamente quanto a dados faltantes ou com valores fora da faixa permitida. Os erros podem então ser comunicados ao respondente, e as respostas, aceitas somente após esses problemas terem sido corrigidos.

CONSIDERAR AFERIÇÕES DIRETAS

Os avanços nos instrumentos de aferição e nos ensaios biológicos estão criando alternativas aos questionários e entrevistas para medir muitas condições e exposições comuns. Por exemplo, a medida direta da atividade física por meio de acelerômetros pequenos produz estimativas mais objetivas e precisas sobre a atividade total, padrões de actigrafia e gasto energético do que seria possível por meio de questionários (9). Sensores acoplados aos sujeitos à noite podem medir de forma mais acurada a quantidade e a qualidade do sono (10). A aferição de níveis séricos de nutrientes como vitamina D fornece uma medida mais acurada da exposição ao nutriente do que perguntar sobre o consumo de alimentos contendo vitamina D. É preciso estar alerta a novas tecnologias, muitas vezes acopladas a dispositivos eletrônicos *wireless*, que medem diretamente características previamente avaliadas apenas indiretamente por meio de questionários e entrevistas.

RESUMO

1. Em grande parte dos estudos clínicos, a qualidade dos resultados depende da qualidade e da adequação dos **questionários** e **entrevistas**. Os investigadores devem buscar garantir o máximo possível de **validade** e **reprodutibilidade** para os **instrumentos** antes de dar início ao estudo.
2. As **questões abertas** permitem que os sujeitos as respondam sem as limitações impostas pelo investigador; e as **questões fechadas** são mais fáceis de serem respondidas e analisadas. As alternativas de respostas para as questões fechadas devem ser **exaustivas** e **mutuamente exclusivas**.
3. As questões devem ser **claras**, **simples**, **neutras** e **apropriadas** para a população que será estudada. Os investigadores devem examinar potenciais questões a partir do ponto de vista dos possíveis participantes, procurando por **termos ambíguos** e possíveis dificuldades de resposta, como **questões que juntem duas perguntas em uma só**, **suposições implícitas** e **alternativas de resposta que não combinem com as perguntas**.

4. Os questionários devem ser de **fácil leitura**, e as questões da entrevista devem ser apropriadas para linguagem oral. A **formatação** deve ser apropriada ao método usado para a entrada eletrônica dos dados, com espaço suficiente para as respostas.
5. Para mensuração de **variáveis abstratas**, como atitudes ou condições de saúde, as questões podem ser combinadas em **escalas de itens múltiplos** que produzam um único escore. Tais escores pressupõem que as questões meçam uma única característica e que as respostas sejam **internamente consistentes**.
6. Os investigadores devem buscar **instrumentos já existentes**, a fim de produzirem resultados válidos e confiáveis. Quando for necessário **modificar medidas existentes** ou **elaborar uma nova medida**, os investigadores devem começar pela coleta de medidas preexistentes que servirão como modelo ou fonte de ideias.
7. O conjunto global de medidas a ser usado em um estudo deve ser **pré-testado** e cronometrado antes do início do estudo. Para instrumentos novos, um pequeno pré-teste inicial pode incrementar a clareza das questões e instruções; mais tarde, estudos-piloto maiores refinarão e testarão a **abrangência**, a **reprodutibilidade** e a **capacidade de detectar mudanças**.
8. **Questionários de autopreenchimento** são mais econômicos do que **entrevistas**, além de serem de fácil padronização e apresentarem um caráter privativo que incrementa a validade das respostas. As **entrevistas**, por sua vez, podem assegurar respostas mais completas e incrementar a validade por propiciar um entendimento mais completo.
9. A aplicação dos instrumentos por **entrevista telefônica assistida por computador, e-mail, dispositivos eletrônicos portáteis** ou por meio da **página do estudo na internet** pode aumentar a eficiência do estudo.

APÊNDICE 15
Exemplo de um questionário sobre fumo

Os itens a seguir foram retirados de um questionário de autopreenchimento em papel usado em nosso Estudo sobre Fraturas Osteoporóticas (*Study of Osteoporotic Fractures*). Observe que as questões ramificadas são seguidas por setas que direcionam o respondente à próxima questão apropriada e que a diagramação é feita de forma organizada, com respostas consistentemente alinhadas à esquerda de cada área. Para acesso à versão *on-line* deste exemplo, veja o *site* www.epibiostat.ucsf.edu/dcr.

1. Ao longo da sua vida, você já fumou pelo menos 100 cigarros?

7. Você já morou na mesma casa, por pelo menos um ano, com alguém que fumava regularmente?

Sim → 8. Por quantos anos, no total, você conviveu com alguém que fumava regularmente?

☐☐ anos

Não

9. Durante todo o tempo que você conviveu com fumantes, quantos cigarros em média essa(s) pessoa(s) fumava(m) enquanto você estava em casa?

☐☐ cigarros por dia

10. Atualmente, você mora na mesma casa que alguém que fuma regularmente?

☐ Sim

☐ Não

11. etc.

REFERÊNCIAS

1. Iarossi G. *The power of survey design: a user guide for managing surveys, interpreting results, and influencing respondents*. Washington, DC: World Bank, 2006. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/6975/350340The0Pow sequence=1, accessed 03/11/13.
2. McDowell I. *Measuring health: a guide to rating scales and questionnaires*, 3rd ed. New York: Oxford University Press, 2006.
3. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*, 4th ed. New York: Oxford University Press, 2009.
4. Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997;314:572.
5. Mangione CM, Berry S, Spritzer K, et al. Identifying the content area for the 51-item National Eye Institute Visual Function Questionnaire: results from focus groups with visually impaired persons. *Arch Ophthalmol* 1998;116:227–233.
6. Mangione CM, Lee PP, Pitts J, et al. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. *Arch Ophthalmol* 1998;116:1496–1504.
7. Kobak KA, Greist JH, Jefferson JW, et al. Computer assessment of depression and anxiety over the phone using interactive voice response. *MD Comput* 1999;16:64–68.
8. Dillman DA, Smyth JD, Christian LM. *Internet, mail, and mixed-mode surveys: the tailored design method*, 3rd ed. Hoboken, NJ: Wiley, 2008.
9. Mackey DC, Manini TM, Schoeller DA, et al. Validation of an armband to measure daily energy expenditure in older adults. *J Gerontol A Biol Sci Med Sci*

2011;66:1108–1113.
10. Girshik J, Fritschi L, Heyworth J, et al. Validation of self-reported sleep against actigraphy. *J Epidemiol* 2012;22:462–468.

1Para formulários *on-line*, a convenção é que as opções mutuamente exclusivas sejam antecedidas por botões de rádio (círculos, também chamados de botões de opção, ou, em inglês, *radio buttons*) e que nas questões tipo “todas as alternativas que se aplicam” sejam utilizadas caixas de seleção (*checkboxes*).

2Os fornecedores comerciais de ferramentas para construção de questionários *on-line* disponíveis nos Estados Unidos levam muito em consideração a legibilidade, em parte porque a Seção 508 do Ato de Reabilitação, de 1973, exige que as agências federais norte-americanas tornem seus formulários eletrônicos acessíveis para pessoas com deficiências. A maioria dos fornecedores comerciais são certificados como aderindo à Seção 508.

CAPÍTULO 16

Gerenciando dados

Michael A. Kohn, Thomas B. Newman e Stephen B. Hulley

Como visto em capítulos anteriores, para realizar uma pesquisa clínica, é necessário definir o delineamento, delimitar a população de estudo e especificar as variáveis preditoras e de desfecho. Ao final do processo, a maior parte das informações sobre os sujeitos e as variáveis estará em um **banco de dados** computadorizado que servirá para armazenar, atualizar e monitorar os dados, bem como para formatá-los para análises estatísticas. Nesse banco de dados, também é possível armazenar **dados administrativos**, como listagens de chamadas telefônicas, agendas de visitas e registros de reembolsos. Bancos de dados simples que consistem em tabelas de dados individuais podem ser mantidos por meio de **planilhas eletrônicas** ou *softwares* estatísticos. Bancos de dados mais complexos que contêm múltiplas tabelas interrelacionadas requerem **sistemas de gerenciamento de bases de dados**.

Para fazer o gerenciamento de dados na pesquisa clínica, é preciso definir as **tabelas de dados**, desenvolver um sistema para a **entrada dos dados** e realizar **consultas** (***queries***) nos dados coletados para fins de **monitoramento** e **análise**. Em ensaios clínicos de grande porte, especialmente naqueles que buscam obter aprovação regulatória de um novo medicamento ou dispositivo, os especialistas que criam os formulários de entradas de dados, gerenciam e monitoram o processo de coleta e formatam e extraem os dados para análise são chamados de **gerentes de dados clínicos** (1). As grandes indústrias farmacêuticas que realizam múltiplos ensaios clínicos dedicam um volume significativo de recursos e de pessoal para o gerenciamento de dados clínicos. Embora geralmente em escala muito menor, os investigadores iniciantes também precisam estar bastante atentos a questões relacionadas ao gerenciamento

de dados.

TABELAS DE DADOS

Todas as bases de dados computadorizadas são compostas por uma ou mais tabelas onde as **linhas** correspondem a **registros** individuais (que podem representar sujeitos, eventos ou transações), e as **colunas** correspondem a **campos** (“atributos” dos registros). Por exemplo, os bancos de dados mais simples consistem em uma tabela única onde cada linha corresponde a um determinado sujeito do estudo e cada coluna corresponde a um atributo específico do sujeito, como nome, data de nascimento, sexo e o valor de uma variável preditora ou de desfecho. Em geral, a primeira coluna corresponde a um **número de identificação único do sujeito** (“IDsujeito”). Usar um identificador único para cada sujeito, sem significado fora do banco de dados, simplifica o processo de desvincular os dados do estudo de identificadores pessoais, o que torna mais fácil assegurar a privacidade dos participantes. Caso o banco de dados contenha tabelas adicionais para dados do exame físico, resultados de exames laboratoriais ou ligações telefônicas, a primeira coluna de cada uma dessas tabelas deve ser um identificador único, como IDexame, IDlab ou IDligacao. O identificador único para uma tabela de dados também é denominado **chave principal** da tabela.

A Figura 16.1 mostra uma tabela de dados simplificada para um estudo de coorte hipotético (inspirado em um estudo real [2]) sobre a associação entre icterícia neonatal e escore de QI aos cinco anos de idade. Cada linha da tabela corresponde a um determinado sujeito do estudo, e cada coluna corresponde a um atributo desse sujeito. A variável preditora dicotômica é se o indivíduo teve ou não “Icterícia”, e a variável de desfecho contínua é “QI”, que corresponde ao escore de QI aos cinco anos de idade.

IDSujeito	PNome	DN	Sexo	Icterícia	DataEx	PesoKg	EstCm	QI
2101	Robert	06/01/2005	M	1	29/01/2010	23,9	118	104
2322	Helen	06/01/2005	F	0	29/01/2010	18,3	109	94
2376	Amy	13/01/2005	F	1	22/03/2010	18,5	117	85
2390	Alejandro	14/01/2005	M	0				
2497	Isiah	18/01/2005	M	0	18/02/2010	20,5	121	74
2569	Joshua	23/01/2005	M	1	13/02/2010	24,8	113	115
2819	Ryan	26/01/2005	M	0				
3019	Morgan	29/01/2005	F	0	09/02/2010	19,1	105	105
3031	Cody	15/02/2005	M	0	16/04/2010	15,2	107	132
3290	Amy	16/02/2005	F	1	12/04/2010	18,0	102	125
3374	Zachary	21/02/2005	M	1				
3625	David	22/02/2005	M	1	10/02/2010	19,2	114	134
3901	Jackson	28/02/2005	M	0				

FIGURA 16.1 Tabela de dados simplificada para um estudo de coorte sobre a associação entre icterícia neonatal e escore de QI aos cinco anos de idade. A variável preditora dicotômica é “Ictericia”, ou seja, se o escore de bilirrubina total chegou a 25 mg/dL ou mais nos primeiros dois dias após o nascimento, e a variável de desfecho contínua é “QI”, o escore de QI do participante aos cinco anos de idade. Os sujeitos 2390, 2819, 3374 e 3901 não foram examinados aos cinco anos de idade.

Se os dados do estudo forem limitados a uma **tabela única**, como a mostrada na Figura 16.1, eles podem ser facilmente acomodados em uma planilha eletrônica ou pacote estatístico. Um banco de dados composto por uma tabela única bidimensional costuma ser denominado **arquivo plano**. Muitos pacotes estatísticos possuem funcionalidades adicionais para acomodar mais de uma tabela, mas, no seu núcleo, a maioria continua se baseando em bancos de dados de arquivos planos.

A necessidade de se incluir mais de uma tabela no banco de dados do estudo (e substituir planilhas eletrônicas ou pacotes estatísticos por programas de gerenciamento de dados) normalmente surge quando o estudo acompanha o registro de múltiplos resultados laboratoriais, medicamentos ou outras aferições repetidas para cada sujeito do estudo. Uma única tabela de dados com uma linha por sujeito do estudo não consegue acomodar um número grande e variável de medidas repetidas. Assim, o banco de dados deve registrar os medicamentos, os resultados de exames e outras medidas repetidas em tabelas separadas, distintas da tabela dos sujeitos do estudo. Cada linha em uma dessas tabelas separadas corresponde a uma medição individual, incluindo, por exemplo, o tipo de

medida, a data/hora em que ela foi feita e seu resultado ou valor. Um dos campos da linha deve incluir o número de identificação do sujeito para poder relacionar a aferição com os campos específicos do sujeito. Nessa **base de dados relacional com múltiplas tabelas**, a relação entre a tabela dos sujeitos e as tabelas das aferições é chamada de **um para muitos**. Estritamente falando, o termo “relacional” tem pouco a ver com as relações entre as tabelas. De fato, “relação” é o termo formal da teoria dos conjuntos matemáticos para uma tabela de dados (3, 4).

Embora os participantes do nosso estudo sobre icterícia tenham feito o exame de QI apenas aos cinco anos, a maioria deles foi também avaliada em outros momentos, durante os quais, junto com outras aferições, foram aferidos o peso e a altura. Os dados sobre altura e peso foram usados para calcular o índice de massa corporal (IMC) e os percentis das curvas de crescimento. (Ver “Extraindo Dados [Consultas]” mais adiante neste capítulo.) A melhor forma de acomodar esses dados é em uma tabela separada de avaliações, na qual cada linha corresponde a uma determinada avaliação, e as colunas representam a data da avaliação, os seus resultados e o número de identificação do sujeito, permitindo fazer a relação com dados da tabela de sujeitos, como sexo, data de nascimento (DN) e ocorrência de icterícia neonatal (Figura 16.2). Nessa estrutura de banco de dados com duas tabelas, para realizar consultas na tabela de avaliações sobre todas as avaliações realizadas em um determinado período de tempo, deve-se fazer a busca na coluna da data em que foram realizadas. Mudanças em campos específicos ao sujeito, como data de nascimento, são feitas em apenas um local, sendo preservada a consistência. Campos envolvendo identificadores pessoais, como nome e data de nascimento, aparecem apenas na tabela de sujeitos. As outras tabelas são relacionadas a essas informações, por meio do número de identificação do sujeito (IDsujeito). O banco de dados pode, ainda , acomodar sujeitos que não fizeram nenhuma avaliação (como, neste exemplo, Alejandro, Ryan, Zachary e Jackson).

Sujeitos: Tabela

IDSujeito	PNome	DN	Sexo	Ictericia
2101	Robert	06/01/2005	M	1
2322	Helen	06/01/2005	F	0
2376	Amy	13/01/2005	F	1
2390	Alejandro			
2497	Isiah			
2569	Joshua			
2819	Ryan			
3019	Morgan			
3031	Cody			
3290	Amy			
3374	Zachary			
3625	David			
3901	Jackson			

Registro: 2 de 13

alia es: Tabela

IDaval	IDSujeito	DataAval	Peso g	stCm
608	2322	29/01/2010	18,3	109
609	2101	29/01/2010	22,0	118
610	2376	01/02/2010	18,3	117
611	3290	05/02/2010	17,6	102
612	3019	09/02/2010	19,1	105
613	3625	10/02/2010	19,2	114
614	2569	13/02/2010	24,8	113
615	2497	18/02/2010	20,5	121
616	3031	26/02/2010	15,5	102
617	2322	19/03/2010	18,6	109
618	2376	22/03/2010	18,5	117
619	3290	26/03/2010	17,8	101
620	2322	05/04/2010	19,1	110
621	3290	12/04/2010	18,0	102
622	3031	16/04/2010	15,2	107
623	3031	05/03/2010	15,6	108

FIGURA 16.2 O banco de dados com duas tabelas do estudo sobre icterícia neonatal é composto por uma tabela dos sujeitos do estudo na qual cada linha corresponde a um único sujeito e uma tabela de avaliações, na qual cada linha corresponde a uma determinada visita de avaliação. Por exemplo, o sujeito 2322 é identificado na primeira tabela como Helen, nascida em 06/01/2005; na segunda tabela anônima, são mostradas três avaliações para esse mesmo sujeito. Uma vez que um sujeito pode realizar múltiplas avaliações, a relação entre as duas tabelas é do tipo "um para muitos". O campo IDsujeito na tabela de avaliações relaciona os dados específicos de cada avaliação com os dados específicos do sujeito.

A manutenção de um registro detalhado dos exames laboratoriais também requer uma tabela separada. A icterícia neonatal é apresentada aqui como um campo dicotômico específico de cada sujeito. Caso os investigadores necessitem de toda a trajetória dos níveis de bilirrubina após o nascimento, o banco de dados deve incluir uma tabela separada contendo os resultados dos exames laboratoriais. Nessa tabela, cada linha corresponde a um resultado de exame, com campos para data/hora da realização do exame, tipo de exame laboratorial (bilirrubina total),

resultado (nível de bilirrubina) e IDsujeito. Esse último campo permite relacionar esses dados com as informações específicas de cada sujeito (Figura 16.3).

Sujeitos: Tabela

ID ujeito	ome	D	e o	Ictericia
2101	Robert	06 01 2005	M	1
2322	Helen	06 01 2005		0
2376	Amy	13 01 2005		1
2390	Alejandro			
2497	Isiah			
2569	Joshua			
2819	Ryan			
3019	Morgan			
3031	Cody			
3290	Amy			
3374	Zachary			
3625	David			
3901	Jackson			

Registro: 3

esult a e: Tabela

IDResult ab	ID ujeito	IDlab	Result ab	Data ab
28	2322	DH	300 0	30 01 2010
37	2376	bili tot	22 3	13 01 2005
38	2376	bili tot	25 1	14 01 2005
39	2376	bili tot	29 4	15 01 2005
40	2376	bili tot	22 1	16 01 2005
41	2376	bili tot	19 0	17 01 2005
42	2390	leuco	14 1	14 01 2005
43	4390	Hg	10 1	14 01 2005
44	2390	HC	32 1	14 01 2005
45	2390	la	403 0	14 01 2005

FIGURA 16.3 Relacionamento entre a tabela de sujeitos e a tabela de resultados de exames laboratoriais. Os resultados de exames laboratoriais capturam a trajetória da bilirrubina total de Amy ao longo dos seus primeiros cinco dias de vida.

Os dados administrativos de um estudo, como listagens de ligações telefônicas, agendas das visitas e registros sobre reembolsos, também requerem múltiplas tabelas separadas. No estudo sobre icterícia neonatal, diversas ligações telefônicas foram feitas aos pais de cada criança. Seria difícil ou até mesmo impossível registrar todas essas ligações em uma única tabela de dados com uma linha por sujeito. Portanto, foi construída uma tabela separada, com uma linha por ligação telefônica, e o campo IDsujeito permitiu relacionar essa tabela com o sujeito sobre o qual foi feita a ligação.

O processo de estruturar um banco de dados com múltiplas tabelas relacionadas, em vez de tentar acomodar os dados em uma única tabela grande e complexa, é denominado **normalização**. Alguns gerentes de dados se referem à normalização como o processo de converter uma ou algumas tabelas "baixas e gordas" em várias tabelas "altas e magras" (1).

A normalização elimina o armazenamento redundante e a possibilidade de inconsistências. Programas para bancos de dados relacionais podem ser programados para assegurar a **integridade referencial**, ou seja, não permitir criar um registro de avaliação, resultado de exame laboratorial ou ligação telefônica para um sujeito que não exista na tabela de sujeitos. Da mesma forma, podem prevenir que um sujeito seja apagado a menos que todas as suas avaliações, resultados de exames e registros de ligações telefônicas sejam também removidos.

Dicionário de dados, tipos de dados e domínios

Até aqui apresentamos as tabelas apenas no modo planilha de dados. Cada coluna ou campo tem um nome e, implicitamente, um tipo de dado e uma definição. Na Tabela “Sujeitos” da Figura 16.2, “PNome” é um campo de formato texto curto que contém o primeiro nome do sujeito; “DN” é um campo de formato data que contém a data de nascimento; e “Icterícia” é um campo do tipo sim/não que indica se os níveis de bilirrubinas foram superiores a 25 mg/dL nos primeiros dois dias após o nascimento. Na Tabela “Exames”, “Peso” é o peso em quilogramas expresso em número real e “QI” é um número inteiro que representa o escore de QI. O **dicionário de dados** deixa explícitas essas definições das colunas. A Figura 16.4 mostra as tabelas de sujeitos e de exames no modo *design* (ou modo de “dicionário de dados”). Note que o dicionário de dados é, na prática, uma tabela, na qual as linhas representam os campos, havendo colunas para nome, tipo e descrição do campo. Como o dicionário de dados é uma tabela de informações sobre o banco de dados em si, ele é muitas vezes chamado de **metadados**. Embora a Figura 16.4 mostre dois dicionários de dados, um para a tabela “Sujeitos” e um para a tabela “Exames”, o banco de dados completo pode ser visualizado como tendo um único dicionário de dados, em vez de um dicionário de dados para cada tabela. Para cada campo no banco de dados, o dicionário de dados único requer a especificação do nome da tabela do campo, além do nome, tipo e descrição do campo e faixa de valores permitidos.

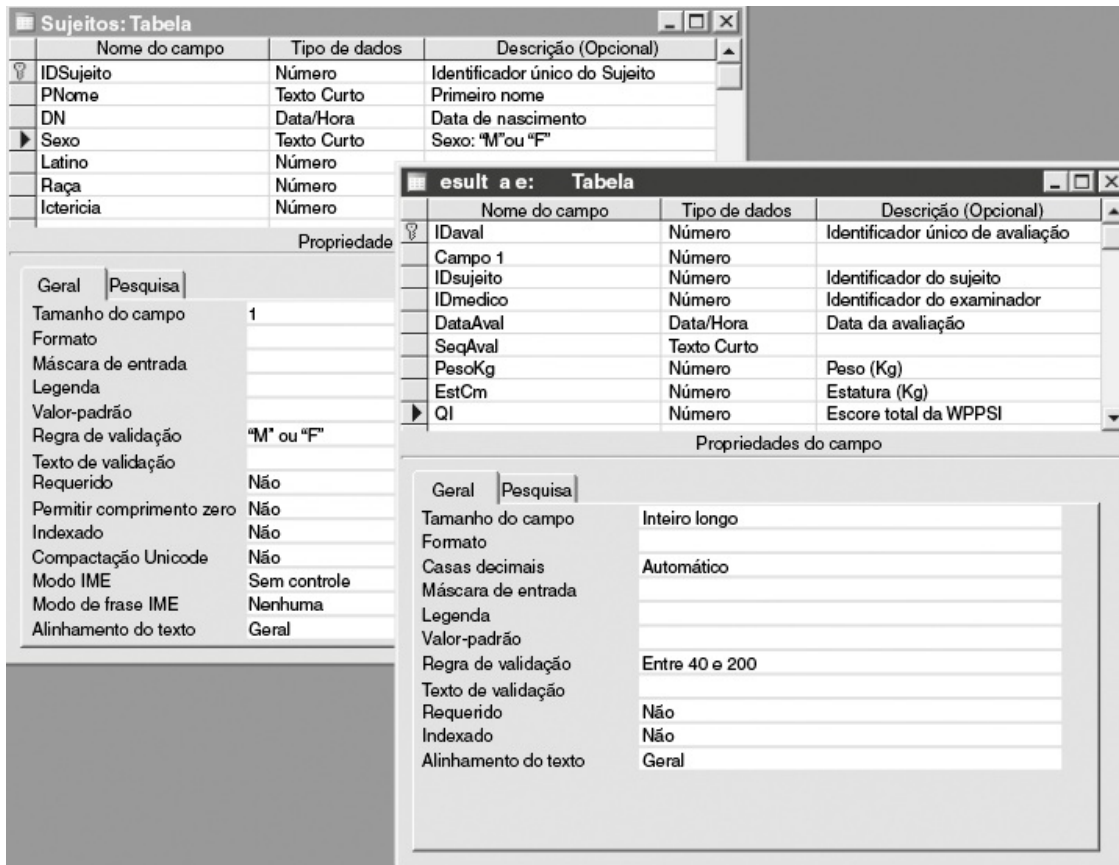



FIGURA 16.4 Tabela dos sujeitos do estudo ("Sujeitos") e tabela das aferições ("Exames") no modo "dicionário de dados". Cada variável ou campo tem um nome, um tipo de dados, uma descrição e um domínio ou conjunto de valores permitidos.

Cada campo possui também um **domínio**, ou seja, uma faixa de valores permitidos. Por exemplo, os valores permitidos para o campo "Sexo" são "M" e "F". O programa não permitirá a entrada de nenhum outro valor nesse campo. Da mesma forma, o campo "QI" só permite números inteiros entre 40 e 200. Os gerentes de dados dos ensaios clínicos geralmente se referem às regras de validação como "verificações de edição" (1). Criar regras de validação que definem os valores permitidos garante algumas proteções contra erros de entrada de dados. Alguns tipos de dados trazem em si regras de validação automáticas. Por exemplo, o programa de gerenciamento de banco de dados sempre rejeitará a data 31 de abril.

Nomes das variáveis

A maioria dos *softwares* de planilhas eletrônicas, análises estatísticas e gerenciamento de bases de dados permite cabeçalhos e nomes de variáveis longos. Há diversas filosofias e convenções para elaboração dos nomes, mas sugerimos que eles sejam suficientemente curtos, para serem fáceis de digitar, e suficientemente longos, para serem autoexplicativos. Embora sejam geralmente permitidos pelo *software*, sugerimos evitar espaços e caracteres especiais nos nomes das variáveis. Nós distinguimos palavras separadas no nome de uma variável por meio de uma técnica denominada "InterCaps", na qual uma palavra nova inicia com letra maiúscula, mas outros pesquisadores preferem usar o caractere sublinhado, às vezes denominado em português pela expressão inglesa *underline* (_). É geralmente preferível que o nome da variável descreva o campo, em vez de sua localização no formulário de coleta de dados (p. ex., "JaFumouCigarros" ou "JaFumou", em vez de "Questao1"). A maioria dos *softwares* permite designar um **rótulo de variável** mais longo, descritivo e fácil de compreender, para ser usado nos formulários de coleta de dados e nos relatórios no lugar do nome compacto da variável.

Elementos comuns para os dados

Diversas organizações de financiamento e regulamentação lançaram iniciativas para desenvolver elementos comuns para as bases de dados de várias áreas específicas da pesquisa clínica. Essas organizações incluem agências governamentais, como o National Institute for Neurological Disorders and Stroke (5), o National Cancer Institute (6), o United States Food and Drug Administration (7) e o European Medicines Agency, assim como associações não governamentais e sem fins lucrativos, como o Clinical Data Interchange Standards Consortium (CDISC) (8).

A justificativa é que os diferentes estudos na mesma área de conhecimento clínico muitas vezes necessitam fazer as mesmas aferições. Padronizar as estruturas dos registros, os nomes/definições dos campos, os tipos/formatos de dados e os formulários de coleta de dados (formulários de relato de caso) elimina o problema de "reinventar a roda", que é frequente nos novos estudos (5), e permite compartilhar e combinar os dados de diferentes estudos. Isso envolve criar um dicionário de dados e um conjunto de instrumentos de coleta de dados com instruções que

todos os investigadores de uma determinada área de conhecimento são estimulados a usar. Para poder ser considerado atualizado academicamente em uma determinada área de pesquisa, o pesquisador deve conhecer os padrões de dados existentes naquela área.

ENTRADA DE DADOS

Todos os bancos de dados, independentemente de consistirem em uma ou várias tabelas, ou de usarem programas de planilhas eletrônicas, de análises estatísticas ou de gerenciamento de bancos de dados, requerem um mecanismo para **popular as tabelas de dados** (isto é, inserir os dados).

Transcrição por teclado

Historicamente, a maneira comum para alimentar um banco de dados tem sido coletar os dados primeiramente em **formulários em papel**. Nos ensaios clínicos, o formulário em papel para coleta de dados relativos a um sujeito específico costuma ser chamado de **formulário de relato de caso** (***case report form,*** **CRF**). O investigador ou um membro da equipe de pesquisa pode preencher o formulário ou, em alguns casos, o próprio participante pode fazê-lo. Um membro da equipe do estudo então pode transcrever os dados do formulário em papel por meio do teclado para as tabelas do computador. A transcrição pode ocorrer diretamente para as tabelas de dados (p. ex., a resposta da questão 3 do sujeito 10 ficaria na casela da linha 10, coluna 3) ou por formulários visualizados na tela do computador projetados para facilitar a entrada dos dados, incluindo mecanismos para validação automática desses dados. A transcrição deve ocorrer tão logo quanto possível após a coleta; assim, o sujeito e o entrevistador, ou coletador dos dados, ainda estarão disponíveis se as respostas estiverem incompletas ou fora da faixa de valores permitidos. Além disso, como discutido mais adiante neste capítulo, o monitoramento de problemas nos dados (p. ex., valores extremos) e as análises preliminares podem ocorrer apenas quando os dados estiverem no banco de dados computadorizado.

Quando se transcreve dados a partir de formulários em papel, pode-se considerar a possibilidade de fazer **entrada dupla de dados** para

assegurar fidelidade na transcrição. O programa de gerenciamento de bancos de dados compara os dois valores entrados para cada variável e apresenta uma lista de valores que não combinam. Entradas discrepantes são então comparadas com os formulários originais e corrigidas. A entrada dupla de dados identifica a entrada incorreta ao custo de duplicar o tempo necessário. Uma alternativa é fazer entrada dupla para uma amostra aleatória dos dados. Se a taxa de erros for aceitavelmente baixa, é improvável que a entrada dupla dos dados remanescentes compense o esforço e o custo.

Entrada distribuída de dados

Se a coleta de dados ocorrer em múltiplas localidades, os formulários em papel poderão ser enviados via correio ou fax para uma central onde serão transcritos para dentro do banco de dados eletrônico, mas isso é cada vez mais raro. O mais comum é que a transcrição dos dados seja feita no próprio local onde foram coletados diretamente na base de dados do estudo por meio de formulários *on-line*. Caso a conexão da internet seja um problema, os dados podem ser armazenados em um computador localizado na sede da coleta e transmitidos on-line ou por meio de um dispositivo de memória portátil, como um *pendrive*. Regras governamentais exigem que informações de saúde em meio eletrônico sejam enviadas sem identificadores ou transmitidas de forma segura (p. ex., criptografadas ou protegidas por senha).

Captura eletrônica de dados

A coleta de dados primários em papel sempre terá seu lugar garantido na pesquisa clínica como uma forma rápida e fácil de capturar dados em um meio não volátil usando caneta e papel. Entretanto, escrever à mão em formulários em papel é cada vez menos comum. Em geral, os estudos devem coletar os dados primariamente por meio de **formulários eletrônicos**. Nos ensaios clínicos, os formulários eletrônicos são denominados **formulários eletrônicos de relato de caso** (***electronic case report forms*, eCRFs**). A entrada de dados por meio de formulários *on-line* tem muitas vantagens:

- Os dados são **digitados diretamente** para as tabelas, sem uma segunda

etapa de transcrição, eliminando essa fonte de erros.

- O formulário eletrônico pode incluir **verificações de validação** e informar imediatamente quando um valor entrado estiver fora da faixa permitida.
- O formulário eletrônico pode também incorporar **lógicas de pulos**. Por exemplo, uma questão sobre número de maços por dia aparecerá apenas se o sujeito responder “sim” para uma questão sobre ser fumante de cigarros.
- O formulário pode ser visualizado, e os dados inseridos em dispositivos **portáteis wireless**, como *tablets* (p. ex., iPad), *smartphones* ou *notebooks*.

Quando usamos formulários on-line para captura eletrônica de dados, eventualmente faz sentido imprimir um registro em papel dos dados imediatamente após a coleta. Essa ação é análoga a imprimir um recibo depois de uma transação em um terminal de autoatendimento bancário. A impressão é uma “fotografia” do registro imediatamente após a coleta de dados e pode ser usada como o original ou como documento fonte, se a versão em papel for exigida.

Respostas codificadas *versus* texto livre

Para definir uma variável ou campo em uma tabela de dados deve-se especificar uma faixa de valores permitidos. Para as análises futuras, é preferível limitar as respostas a uma faixa de valores codificados do que permitir respostas de texto livre. Essa distinção é a mesma que aquela feita no Capítulo 15 sobre questões “fechadas” *versus* questões “abertas”. Se não houver clareza sobre qual faixa de respostas possíveis usar, uma coleta inicial de dados durante o pré-teste poderá permitir respostas de texto livre que serão subsequentemente usadas para desenvolver opções de respostas codificadas.

As opções de respostas para uma questão devem ser **exaustivas** (todas as opções possíveis são disponibilizadas) e **mutuamente exclusivas** (não pode haver duas respostas corretas). Um conjunto de opções de respostas mutuamente exclusivas pode sempre ser tornado coletivamente exaustivo adicionando-se uma resposta do tipo “outro”. Formulários de coleta de

dados *on-line* oferecem três formatos possíveis para exibir opções de respostas mutuamente exclusivas e exaustivas: listas do tipo *drop-down*, listas de escolha (lista de campos) ou grupos de opções (Figura 16.5). Esses formatos são familiares a qualquer sujeito de estudo ou pessoa do sistema de entrada de dados que tenha tido contato com formulários *on-line*. É importante ter em mente que uma lista *drop-down* economiza espaço na tela, mas não funcionará bem caso precise ser impressa para coleta de dados via papel, uma vez que as opções de resposta não estarão visíveis.

FIGURA 16.5 Formatos para entrada de dados a partir de uma lista de respostas mutuamente exclusiva e coletivamente exaustiva. A lista *drop-down* (A, sendo o menu *drop-down* mostrado no painel inferior) economiza espaço na tela, mas não é adequada quando o formulário da tela será impresso em papel para a coleta de dados. Tanto a lista de escolhas (que na verdade é um menu *drop-down* formatado de forma a ser visualizado sempre; B) quanto o grupo de opções (C) ocupam mais espaço na tela, porém são mais adequados para impressão.

Uma questão com um conjunto de respostas mutuamente exclusivas corresponde a um único campo na tabela de dados. Por outro lado, as respostas a uma questão do tipo “Todas as alternativas que se aplicam” não são mutuamente exclusivas e correspondem a tantos campos sim/não quanto forem as opções disponíveis. Por convenção, opções de respostas para questões do tipo “Todas as alternativas que se aplicam” utilizam caixas de seleção (*checkboxes*) retangulares no lugar dos botões de rádio redondos usados para grupos de opções com respostas mutuamente exclusivas. Como discutido no Capítulo 15, desaconselhamos o uso de questões do tipo “Todas as alternativas que se aplicam” e preferimos que seja solicitada uma resposta sim/não para cada item, pois, caso contrário, uma resposta não marcada poderia tanto significar “não se aplica” quanto “não respondido”. Ao codificar variáveis sim/não (dicotômicas), recomendamos que o número 0 represente *não* ou *ausente* e + represente *sim* ou *presente*. Com essa codificação, o valor médio da variável é interpretável como a proporção do atributo.

Importando medições e resultados laboratoriais

Muitos dados do estudo, como informações demográficas da linha de base

nos registros hospitalares, resultados laboratoriais no sistema de informática do laboratório e medições feitas por absorciometria de dupla emissão de raios X (DEXA) e monitores Holter, já estão em formato eletrônico. Quando possível, esses dados devem ser importados diretamente no banco de dados do estudo para evitar o trabalho e os potenciais erros de transcrição envolvidos ao entrar os dados uma segunda vez. Por exemplo, no estudo sobre icterícia neonatal, os dados demográficos e as informações para contato foram obtidas da base de dados do hospital. Sistemas informatizados quase sempre produzem arquivos com delimitadores de texto ou com caracteres de largura de coluna fixa (*fixed-column-width characters*) que podem ser importados pelo sistema de banco de dados. Em ensaios clínicos, esse tipo de informações enviadas em lotes (*batch-uploaded*) é denominado dado não CRF, em referência ao termo em inglês *case report form* (formulário de relato de caso) (1).

Softwares para gerenciamento de dados

Agora que já discutimos tabelas de dados e entrada de dados, podemos fazer a distinção entre o *front-end* e o *back-end* do banco de dados do estudo. O ***back-end*** consiste nas tabelas de dados em si. O ***front-end*** ou "interface" consiste nos formulários *on-line* usados para entrada, visualização e edição dos dados. A Tabela 16.1 lista alguns dos aplicativos usados no gerenciamento de dados de pesquisas clínicas. Bancos de dados simples que consistem em uma tabela única podem usar programas de planilhas eletrônicas ou pacotes estatísticos para a tabela de dados *back-end*. Nesse caso, o pessoal do estudo entra os dados diretamente nas caselas, evitando o uso de formulários *front-end*. Bancos de dados mais complexos consistem em múltiplas tabelas e requerem um programa de **banco de dados relacional** para manter as tabelas *back-end*. Se os dados forem coletados primeiramente em formulários em papel, para inseri-los será necessário transcrevê-los em formulários *on-line*.

TABELA 16.1 Alguns aplicativos utilizados no gerenciamento de dados de pesquisas

Planilhas Eletrônicas
Microsoft Excel
Planilhas do Google Drive*
Apache OpenOffice Calc*

Análises Estatísticas

- Statistical Analysis System (SAS)
- Statistical Package for the Social Sciences (SPSS)
- Stata
- R*
- EpiInfo* (apenas para Windows)

Sistemas Integrados de Bancos de Dados

- Microsoft Access (apenas para Windows)
- Filemaker Pro

Sistemas de Bases de Dados Relacionais

- Oracle
- SQL Server
- MySQL*
- PostgreSQL*

Plataformas *On-line* Integradas para Gerenciamento de Dados de Pesquisas

- Research Electronic Data Capture* (REDCap – uso acadêmico apenas, hospedado pela instituição do investigador)
- QuesGen (uso primariamente acadêmico, hospedado pela empresa que vende o produto)
- MediData RAVE (uso primariamente corporativo e não acadêmico, hospedado pela empresa que vende o produto)
- Oracle InForm (uso corporativo não acadêmico, hospedado pela companhia)
- Datalabs EDC (corporativo, hospedado pela empresa que vende o produto)
- OnCore
- OpenClinica

Ferramentas para Construir Questionários On-line

- SurveyMonkey
- Zoomerang
- Qualtrics

* Gratuitos

Como discutido no Capítulo 15, diversas ferramentas, incluindo o SurveyMonkey, o Zoomerang e o Qualtrics, estão disponíveis para desenvolver questionários *on-line* para serem enviados por *e-mail* aos participantes do estudo ou postados na página do estudo na internet. Todas essas ferramentas oferecem opções para formatar questões de múltipla escolha, incorporar a lógica de pulos, bem como agregar, relatar e exportar os resultados.

Alguns **pacotes estatísticos**, como o SAS, disponibilizam módulos de entrada de dados. Programas de **bancos de dados integrados,** como o Microsoft Access e o Filemaker Pro, também fornecem inúmeras ferramentas para o desenvolvimento de formulários na tela.

Cada vez mais os estudos utilizam plataformas integradas *on-line* para

gerenciamento de dados de pesquisa. O **REDCap** (Research Electronic Data Capture) é um sistema de coleta de dados desse tipo desenvolvido por um consórcio acadêmico liderado pela Universidade de Vanderbilt. Ele permite aos pesquisadores construir formulários de entrada de dados, inquéritos e inquéritos com formulários anexados para entrada de dados. O REDCap é disponibilizado apenas para pesquisadores acadêmicos e deve estar hospedado na instituição do investigador. Trata-se de uma excelente ferramenta tipo “faça você mesmo” para pesquisadores iniciantes que permite desenvolver, em pouco tempo, questionários e formulários para coleta de dados na tela. Também fornece acesso a um repositório de instrumentos de coleta de dados que podem ser baixados no computador. Assim como ocorre com qualquer ferramenta “faça você mesmo” para desenvolvimento na internet, as opções de customização e as funcionalidades avançadas são limitadas. Um banco de dados desenvolvido no REDCap consiste em uma única tabela contendo uma linha para cada um de um número fixo de “eventos” definido pelo usuário para cada sujeito do estudo. Ele não permite acompanhar detalhadamente um número grande e variável de medidas repetidas por sujeito do estudo, como resultados de exames laboratoriais, sinais vitais, medicamentos ou registros de ligações telefônicas. O REDCap também não é capaz de realizar validações mais sofisticadas dos dados, consultas (ver adiante neste capítulo) ou elaborar relatórios, mas é fácil exportar os dados dessa plataforma para os pacotes estatísticos.

Plataformas mais poderosas para gerenciamento de dados de pesquisa *on-line*, como o **QuesGen**, o **MediData RAVE** ou o **Oracle InForm**, podem acomodar estruturas de dados complexas, além de oferecer ferramentas mais sofisticadas para validação, consultas e relato dos dados. As empresas que oferecem acesso a essas plataformas também fornecem suporte técnico e auxílio na instalação. Embora possa haver algum custo adicional envolvido, essas soluções valem a pena quando as ferramentas tipo “faça você mesmo” carecem de sofisticação suficiente para atender às necessidades do estudo.

EXTRAINDO DADOS (CONSULTAS)

Após o banco ter sido criado e os dados terem sido coletados, o

investigador irá querer **organizar, classificar, filtrar** e **ver** (“consultar”) os dados que coletou. **Consultas** são usadas para monitorar a entrada de dados, relatar o andamento do estudo e, finalmente, analisar os resultados. A linguagem-padrão para manipular dados em um **banco relacional** é denominada **SQL** (*Structured Query Language*, Linguagem de Consulta Estruturada). Todos os sistemas de bancos de dados relacionais usam uma ou outra variante de SQL, mas a maioria fornece uma interface gráfica para a construção de consultas sem a necessidade de aprendizado da linguagem pelos pesquisadores clínicos.

Uma consulta pode **unir** dados de duas ou mais tabelas, mostrar apenas campos selecionados e filtrar os dados para mostrar apenas os registros que atendem certos critérios. A consulta pode também calcular valores a partir dos dados brutos das tabelas. A Figura 16.6 mostra o resultado de uma consulta no banco de dados de icterícia neonatal; foram usados filtros para incluir apenas os meninos examinados em fevereiro e calculou-se a idade em meses (a partir da data de nascimento até a data de exame) e o IMC, a partir do peso e da altura. A consulta também utilizou uma função sofisticada de busca na tabela para calcular os valores dos percentis na curva de crescimento para o IMC da criança. Note que os resultados de uma consulta que une duas tabelas, mostrando apenas alguns campos, selecionando linhas com base em critérios especiais e calculando certos valores, ainda se parecem com uma tabela no modo planilha eletrônica. Um dos princípios do modelo relacional é que operações sobre tabelas geram resultados semelhantes a tabelas. Os dados da Figura 16.6 são facilmente exportados para um pacote estatístico. Note que nenhum identificador pessoal está incluído na consulta.

Meninos Examinados em Fevereiro (Com Percentis do IMC) Selecionar Consulta

IDSujeito	Sexo	DataExame	IdadeMeses	PesoKg	EstCm	IMCcalc	IMCperc
2497	M	18/02/2010	61	20,5	121	14,0	8
2569	M	13/02/2010	60	24,8	113	19,4	99
3031	M	26/02/2010	59	15,5	102	14,9	33
3625	M	10/02/2010	59	19,2	114	14,7	26
4430	M	23/02/2010	59	35	100	35,0	100
5305	M	23/02/2010	60	20,5	116	15,2	43
5310	M	24/02/2010	60	19,6	115	14,8	28

Registro: 5 de 7

FIGURA 16.6 Consulta no modo folha de dados, filtrando para meninos examinados em fevereiro e calculando a idade em meses (a partir da data de nascimento até a data do exame) e o índice de massa corporal (IMC) a partir do peso e da altura. A consulta também utiliza uma função sofisticada de pesquisa em tabela para calcular os valores dos percentis das curvas de crescimento para o IMC da criança. Para o IDsujeito 4430, devido ao fato de ele estar no percentil 100 e ao valor elevado do IMC (35,0), deve-se averiguar se houve erro de entrada de dados.

Identificando e corrigindo erros nos dados

O primeiro passo para evitar erros nos dados é testar a coleta e o sistema de gerenciamento de dados como parte do processo global de pré-testagem do estudo. Todo o sistema (tabela de dados, formulários de entrada de dados e consultas) deve ser testado com dados fictícios. Para ensaios clínicos visando à submissão de dados para o FDA, essa é uma exigência regulatória prevista no Código de Regulamentações Federais (*Code of Federal Regulations*), Capítulo 21, Parte 11 (21 CFR 11) (9).

Já foram discutidas neste capítulo formas de assegurar a fidelidade em transcrições por meio do teclado ou de captura eletrônica após o início da fase de coleta. Assim, valores fora da faixa permitida não deveriam sobreviver ao processo de entrada de dados. Entretanto, é também importante realizar consultas para valores faltantes e *outliers* (i. e., valores extremos que, todavia, estão dentro da faixa de valores permitidos). Por exemplo, um peso de 35 kg é um dado aceitável para uma criança de cinco anos, mas se esse peso for 5 kg maior do que qualquer outro peso no conjunto de dados, ele merece investigação. Muitos sistemas de entrada de dados são incapazes de fazer uma validação com cruzamento de campos, ou seja, os dados das tabelas podem possuir valores dentro da faixa aceitável, porém inconsistentes uns com os outros. Por exemplo, é altamente improvável para uma criança de cinco anos que pesa 30 kg ter uma altura de 100 cm. Embora os valores de peso e altura estejam dentro da faixa permitida, a altura (extremamente baixa para uma criança de cinco anos) está inconsistente com o peso (extremamente alto para uma criança de cinco anos). Inconsistências como essa podem ser suspeitadas por meio de consultas como aquela mostrada na Figura 16.6.

Valores faltantes, *outliers*, inconsistências e outros problemas com os dados são identificados por meio de consultas, sendo então comunicados à equipe do estudo, que pode responder verificando documentos-fonte

originais, entrevistando o participante ou repetindo as aferições. Se o estudo utilizar documentos em papel, quaisquer mudanças nos dados devem ser assinaladas (p. ex., com tinta vermelha), datadas e assinadas. Como discutido mais adiante neste capítulo, bancos de dados eletrônicos devem manter um registro de auditoria sobre todas as mudanças nos dados.

Se os dados forem coletados por vários investigadores de diferentes localidades, as médias e medianas devem ser comparadas entre os investigadores e entre os centros de estudo. Diferenças substanciais entre os investigadores ou centros de estudo indicam diferenças sistemáticas nas medições ou na coleta de dados.

A edição e a limpeza dos dados devem dar maior prioridade a variáveis mais importantes. Por exemplo, em um ensaio clínico randomizado, a variável mais importante é o desfecho, e, portanto, para ela, devem-se minimizar dados faltantes e erros. Por outro lado, erros em outras variáveis, como a data de uma visita, podem não afetar substancialmente o resultado das análises. A edição de dados é um processo iterativo; após os erros serem identificados e corrigidos, os procedimentos de edição devem ser repetidos até que um número muito pequeno de erros importantes seja identificado. Nesse ponto, para alguns estudos, o banco de dados editado é declarado final ou **congelado**, não sendo permitidas futuras modificações (1).

ANÁLISE DOS DADOS

Ao se analisarem os dados, geralmente se criam variáveis novas derivadas a partir dos valores brutos do banco de dados. Por exemplo, variáveis contínuas podem ser dicotomizadas (p. ex., IMC > 25 sendo definido como sobrepeso), novas categorias podem ser criadas (fármacos específicos agrupados como antibióticos) e valores podem ser calculados (*número de anos como fumante* × *número de maços de cigarro por dia* = *anos-maços*). Os dados faltantes devem ser manejados de forma consistente. A resposta “não sei” pode ser recodificada como uma categoria especial, combinada com a resposta “não”, ou então excluída como faltante. Se o estudo estiver usando programas para gerenciamento de bancos de dados, pode-se realizar consultas para derivar novas

variáveis antes de exportá-las para um pacote estatístico. Isso é especialmente importante para variáveis, como os percentis da Figura 16.6 que requerem programação complexa ou uma tabela separada para “pesquisa”. Alternativamente, os novos campos podem ser derivados já no próprio pacote estatístico.

Muitos pesquisadores têm maior familiaridade com pacotes de análise estatística do que com programas de gerenciamento de bases de dados e, portanto, preferem calcular as variáveis derivadas após a exportação.

CONFIDENCIALIDADE E SEGURANÇA

Caso os participantes da pesquisa também sejam pacientes do ambulatório ou hospital do estudo, as informações que os identificam estão protegidas sob o Privacy Rule do Health Insurance Portability and Accountability Act (**HIPAA**) (10); mesmo assim, independentemente de eles serem ou não pacientes, o investigador tem o dever ético e legal de proteger sua confidencialidade. O banco de dados deve atribuir a cada um dos participantes um identificador único (IDsujeito) que não possua significado externo ao banco de dados do estudo (isto é, o IDsujeito não deve incorporar seu nome, iniciais, data de nascimento ou número do prontuário). Todos os campos do banco de dados que contêm informações pessoais devem ser apagados antes de os dados serem compartilhados. Se o banco de dados usar múltiplas tabelas, os identificadores pessoais poderão ser mantidos em uma tabela separada. Bancos de dados que contenham identificadores pessoais devem ser mantidos em servidores seguros acessíveis apenas para membros autorizados da equipe de pesquisa, onde cada um desses membros terá seu próprio nome de usuário e senha. Plataformas *on-line* para gerenciamento de dados, como o REDCap e o QuesGen, permitem designar campos contendo identificadores pessoais dos participantes. Diferentes perfis de usuários podem permitir ou proibir a exportação, alteração ou visualização desses campos.

O sistema de banco de dados deve **auditar toda entrada e edição de dados**. A auditoria permite determinar quando algum elemento dos dados foi modificado, quem o modificou e qual modificação foi feita. Para ensaios clínicos sobre novos medicamentos, essa é uma exigência

regulatória (9). Plataformas de pesquisa *on-line*, como o REDCap, o QuesGen e o MediData RAVE, permitem validação e auditoria automáticas.

Regularmente, deve-se **fazer uma cópia de segurança** do banco de dados do estudo e **armazená-la fora de sua sede**. Periodicamente, o procedimento de criação de cópia de segurança deve ser testado, restaurando a cópia dos dados. Assim como ocorre com a validação e auditoria, plataformas *on-line* como o REDCap, o QuesGen e o MediData RAVE realizam automaticamente cópias de segurança e oferecem mecanismos para assegurar a segurança dos dados. Ao final do estudo, os dados originais, o dicionário de dados, o banco de dados final e as análises do estudo devem ser **arquivados** para uso futuro. Esses arquivos podem ser revistos futuramente, permitindo ao investigador responder a questões sobre a integridade dos dados ou das análises, realizar novas análises para responder a novas questões de pesquisa e compartilhar os dados com outros investigadores.

RESUMO

1. O **banco de dados** do estudo consiste em uma ou mais tabelas de dados onde as **linhas** correspondem aos **registros** (p. ex., sujeitos do estudo), e as **colunas** correspondem a **campos** (atributos dos registros).
2. Identificar os participantes do estudo com um código de identificação (**IDsujeito**) que não tenha significado externo ao estudo permite desvincular os dados do estudo de identificadores pessoais, assegurando a **confidencialidade** dos participantes. Os bancos de dados que possuem **identificadores pessoais** devem ser armazenados em servidores seguros, com acesso restrito e auditorias.
3. Para acomodar um número variável de **medidas repetidas** por sujeito do estudo, como resultados de exames laboratoriais ou medicamentos em uso, devem-se **normalizar** os dados aferidos em tabelas separadas, nas quais cada linha corresponde a uma **aferição**, e não a um sujeito do estudo.
4. O banco de dados do estudo deve também armazenar **dados administrativos**, como **listagens de ligações telefônicas**, **agendas de**

exames e **registros de reembolsos**.

5. O **dicionário de dados** especifica o **nome**, o **tipo de dados**, a **descrição** e a **faixa de valores permitidos** para todos os campos no banco de dados.
6. O **sistema de entrada de dados** é o meio pelo qual as tabelas de dados são alimentadas; a **captura eletrônica de dados** por meio de formulários *on-line* está substituindo a transcrição a partir de formulários em papel.
7. Uma **planilha eletrônica** ou **pacote estatístico** são adequados apenas para os bancos de dados mais simples; para bases de dados complexas, é necessário criar um **banco de dados relacional** usando **programas de gerenciamento de banco de dados**, baseados na linguagem **SQL** (*Structured Query Language*).
8. **Consultas a bancos de dados** ordenam e filtram os dados assim como calculam valores baseados nos valores dos campos. Consultas são usadas para **monitorar a entrada de dados, relatar progressos do estudo** e **formatar os resultados** para análise.
9. A perda do banco de dados pode ser prevenida pela criação regular de **cópias de segurança** (***backups***)**, armazenamentos fora da sede**, e pelo **arquivamento** de cópias de versões-chave do banco de dados para uso futuro.

REFERÊNCIAS

1. Prokscha S. *Practical guide to clinical data management*, 3rd ed. Boca Raton: CRC Press, 2012.
2. Newman TB, Liljestrand P, Jeremy RJ, et al. Outcomes among newborns with total serum bilirubin levels of 25 mg per deciliter or more. *N Engl J Med* 2006;354(18):1889–1900.
3. Codd EF. A relational model of data for large shared data banks. *Communications of the ACM* 1970;13(6):377–387.
4. Date CJ. *An introduction to database systems*, 7th ed. Reading, Mass: Addison-Wesley, 2000.
5. Grinnon ST, Miller K, Marler JR, et al. National Institute of Neurological Disorders and Stroke common data element project—approach and methods. *Clin Trials* 2012;9(3):322–329.
6. NCI. *The National Cancer Institute Cancer Data Standards Registry and Repository*. 2012. Available from: https://cabig.nci.nih.gov/concepts/caDSR/, accessed 9/29/12.

7. FDA. Driving biomedical innovation: initiatives to improve products for patients. October, 2011. Available from: http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM274 accessed 1/29/13.
8. CDISC. The Clinical Data Interchange Standards Consortium Study data tabulation model. 2012. Available from: http://www.cdisc.org/sdtm, accessed 1/29/2013.
9. DHHS. Guidance for industry: computerized systems used in clinical trials. May, 2007. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/G accessed 1/29/2013.
10. DHHS. Protecting personal health information in research: understanding the HIPAA Privacy Rule. 2003. Available from: http://privacyruleandresearch.nih.gov/pr_02.asp, accessed 1/29/2013.

CAPÍTULO 17

Implementando o estudo e controlando a qualidade

Deborah G. Grady e Stephen B. Hulley

A maior parte deste livro concentrou-se no lado esquerdo do modelo de pesquisa clínica, abordando aspectos do delineamento (Figura 17.1). Neste capítulo, enfocaremos o lado direito, abordando aspectos da **implementação**. Mesmo os melhores projetos de pesquisa, cuidadosamente delineados no papel, podem, na prática, funcionar de modo muito diferente do esperado. Pode não haver pessoal técnico disponível com a capacitação adequada, a área física pode não ser a ideal, os participantes podem estar menos dispostos a ingressar no estudo do que o antecipado, a intervenção pode ser mal tolerada e as aferições podem ser de difícil execução. As conclusões de um estudo bem delineado podem ser comprometidas pela ignorância, falta de cuidado, falta de treinamento e de padronização e por outros problemas na finalização e na implementação do protocolo.

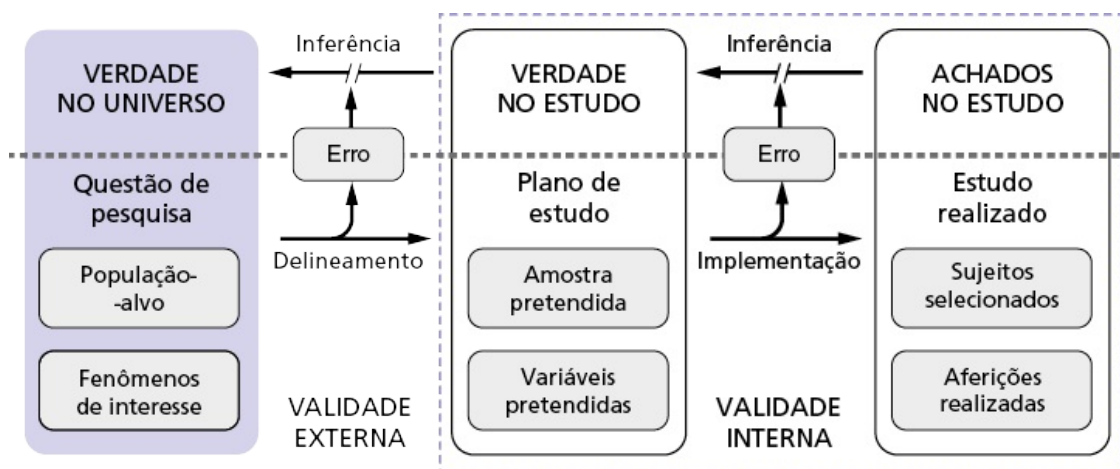



FIGURA 17.1 Este capítulo tem como foco a área dentro da linha verde tracejada: a implementação de um projeto de pesquisa.

A implementação bem-sucedida inicia pela **montagem dos recursos para dar início ao estudo,** que incluem área física e pessoal de pesquisa e financeiro. A tarefa seguinte é **finalizar o protocolo** com um processo de **pré-testagem** dos planos de recrutamento, aferições e intervenções, em uma tentativa de evitar a necessidade de **revisões no protocolo** após o início da coleta de dados. O estudo é então conduzido de forma sistematizada quanto ao **controle de qualidade** de **procedimentos clínicos e laboratoriais** e de **gerenciamento de dados**, seguindo os princípios das **Boas Práticas Clínicas** (GPC, *Good Clinical Practice*).

Algumas das estratégias descritas neste capítulo são mais aplicáveis a estudos de grande porte, com equipes numerosas de pesquisadores distribuídos em múltiplos centros, liderados por investigadores experientes. Entretanto, também são relevantes para investigadores iniciantes que podem ser coinvestigadores nesses estudos de grande porte ou investigadores principais em um estudo menor.

MONTANDO OS RECURSOS

Área física

É possível conduzir alguns estudos inteiramente na internet, usando sistemas interativos *on-line*, intervenções enviadas pelo correio (como medicamentos ou dispositivos), monitoramento remoto, visitas domiciliares para realizar as aferições e entrada *on-line* de dados. Entretanto, a maioria dos estudos ainda requer uma área física para realizar as visitas e as aferições. Esse espaço deve ser de fácil acesso, atraente e de tamanho adequado. Quando não se consegue negociar um espaço adequado cedo no planejamento da pesquisa, pode haver maior dificuldade no arrolamento de participantes, baixa adesão às visitas do estudo, dados incompletos e equipe desmotivada. A **área física** deve ser de fácil acesso aos participantes e dispor de local adequado para estacionamento. Deve também ser acolhedora, confortável e suficientemente espaçosa para acomodar o pessoal de pesquisa e o equipamento para as medições, bem como para armazenar os fármacos do

estudo e os registros relacionados à pesquisa. Se o estudo envolver exame físico dos participantes, o local deverá assegurar privacidade e condições para lavagem das mãos. Se os participantes precisarem ir a outros locais para realizar exames (como o laboratório do hospital ou o serviço de radiologia), estes deverão ser de fácil acesso. Em alguns estudos, como naqueles que envolvem pacientes enfermos ou que são submetidos a intervenções potencialmente danosas, pode ser necessário acesso a equipes e equipamentos de reanimação cardiorrespiratória.

Muitos centros médicos acadêmicos dispõem de **unidades de pesquisa clínica** que fornecem salas totalmente equipadas para pesquisa, com pessoal técnico experiente. As unidades de pesquisa clínica muitas vezes têm condições de realizar exames especializados (como ingesta calórica, densidade óssea e estudos de *clamp*) e podem fornecer acesso a outros serviços (como serviços de recrutamento de participantes, gerenciamento de dados e análise estatística). Essas unidades de pesquisa são uma excelente opção para realizar pesquisas clínicas e translacionais, mas geralmente exigem procedimentos adicionais de inscrição e de avaliação de protocolos, bem como reembolso pelos serviços.

A equipe de pesquisa

As equipes de pesquisa variam em tamanho desde pequenas – às vezes contando apenas com um investigador e um assistente de pesquisa em tempo parcial – até equipes grandes com vários funcionários em dedicação exclusiva para estudos de grande porte. Independentemente do tamanho, todas as equipes de pesquisa devem realizar atividades semelhantes e preencher funções parecidas, descritas na Tabela 17.1. Muitas vezes, uma única pessoa realiza várias dessas atividades. No entanto, muitas funções requerem habilidades especiais, como programação e análises estatísticas. Alguns membros da equipe, como os responsáveis pela parte financeira e pelos recursos humanos, são geralmente empregados pela universidade ou pelo centro médico e cedidos pelo departamento ou unidade do investigador. Independentemente do tamanho da equipe de pesquisa, o **investigador principal (IP)** deve assegurar que cada uma das funções listadas na Tabela 17.1 seja realizada.

TABELA 17.1 Funções dos membros de uma equipe de pesquisa*

CARGO	FUNÇÃO	COMENTÁRIO
Investigador principal	Responsável, em última instância, pelo delineamento, pelo financiamento, pela formação da equipe, pela condução e qualidade do estudo, bem como pelo relato dos resultados	
Diretor de projeto/coordenador da clínica	Responsável pelo gerenciamento da rotina diária de todas as atividades relacionadas ao estudo	Experiente, responsável, meticuloso, com fortes habilidades interpessoais e organizacionais
Responsável pelo recrutamento	Assegura que o número desejado de participantes elegíveis seja arrolado	Detém conhecimento e experiência em lidar com uma série de técnicas de recrutamento
Assistente de pesquisa/funcionário do centro clínico	Realiza os procedimentos para as visitas do estudo e as medições	O exame físico ou outros procedimentos especializados podem requerer licenças ou certificados especiais
Coordenador de controle de qualidade	Assegura que toda a equipe do estudo siga procedimentos operacionais-padrão (POPs) e supervisiona o controle de qualidade	Observa os procedimentos do estudo para garantir a adesão aos POPs, pode supervisionar a auditoria por grupos externos, como o FDA
Gerente de dados	Planeja, testa e implementa a entrada, a edição e o sistema de armazenamento de dados	
Programador/analista	Produz relatórios que descrevem o recrutamento, a adesão e o controle de qualidade e realiza a análises de dados	Trabalha sob a supervisão do investigador principal e do estatístico
Estatístico	Colabora no delineamento do estudo, estima o tamanho de amostra e o poder estatístico, elabora o plano de análises e as diretrizes para o monitoramento dos dados e da segurança e interpreta os achados	Muitas vezes desempenha um papel importante no delineamento, na implementação do estudo, no monitoramento interino, nas análises dos dados e na apresentação dos resultados
Assistente administrativo	Dá apoio secretarial e administrativo, agenda as reuniões, etc.	
Administrador financeiro	Prepara o orçamento e administra os gastos	Prepara projeções para ajudar na gerência do orçamento
Administrador de recursos humanos	Auxilia na preparação das descrições de cargos e funções e na contratação e avaliação de pessoal	Auxilia na administração de questões e problemas relacionados à equipe de pesquisa

* Em estudos de pequeno porte, uma mesma pessoa pode assumir várias dessas funções; alguns membros, como os gerentes financeiros e de recursos humanos, podem ser disponibilizados pelo departamento do investigador e compartilhados com outros

pesquisadores.

Após decidir o número de membros na equipe e a distribuição das tarefas, o próximo passo é trabalhar com um administrador do departamento para encontrar **candidatos** qualificados e experientes. Isso pode ser difícil, pois o treinamento formal para alguns membros da equipe varia, e as exigências do serviço variam entre diferentes estudos. Por exemplo, a posição crucial de diretor de projeto pode ser preenchida por um profissional com formação em enfermagem, farmácia, saúde pública, análises clínicas ou pesquisa farmacêutica, e as tarefas para esse cargo podem variar imensamente.

A maioria das universidades e centros médicos dispõem de métodos formais para divulgar editais para preenchimento de vagas, mas outras formas de divulgação, como jornais e internet, podem ser úteis. A forma mais segura é procurar pessoas de competência comprovada, por exemplo, alguém que trabalhou para um colega seu cujo projeto já esteja concluído.

Também é comum negociar com outros pesquisadores da mesma instituição para contratar integrantes de sua equipe em tempo parcial. Alguns centros médicos acadêmicos ou unidades dentro do centro médico têm à disposição um grupo de coordenadores de pesquisa experientes e outros profissionais que podem ser contratados em tempo parcial.

Liderança e formação da equipe

A qualidade de um estudo conduzido por mais de uma pessoa começa pela integridade e pela capacidade de **liderança do IP**. O investigador principal deve assegurar que toda a equipe esteja adequadamente treinada e certificada para realizar suas tarefas. Ele deve deixar clara a mensagem de que a proteção dos sujeitos da pesquisa, a manutenção da privacidade, a completude e acurácia dos dados e a apresentação adequada dos achados são de fundamental importância. Ele não tem como monitorar cada medição feita por um colega da equipe, mas, se construir uma noção de que está a par de todas as atividades relacionadas ao estudo e de que se importa muito com a proteção dos sujeitos e com a qualidade dos dados, o resto da equipe irá agir da mesma forma. É útil se reunir periodicamente com cada membro da equipe, mostrando que aprecia o seu trabalho e discutindo eventuais problemas e soluções. Um bom líder sabe delegar

responsabilidades e, ao mesmo tempo, constrói um sistema hierarquizado de supervisão que assegura um controle adequado de todos os aspectos do estudo.

Desde o início da fase de planejamento, o investigador deve realizar **reuniões de equipe** regulares com todos os membros da equipe de pesquisa. As reuniões devem ter uma pauta distribuída com antecedência, com relatórios sobre o andamento elaborados pelas pessoas que ficaram responsáveis por áreas específicas do estudo. Essas reuniões fornecem uma oportunidade para descobrir e solucionar problemas e para envolver toda a equipe no processo de desenvolvimento do projeto e na condução da pesquisa. Elas podem ser aprimoradas incluindo-se discussões científicas e atualizações relacionadas ao tema do projeto. Reuniões de equipe regulares são uma excelente fonte de estímulo e de interesse no alcance das metas do estudo e fornecem a oportunidade de educação e treinamento "em serviço".

A maior parte das universidades e centros médicos voltados para pesquisa oferece uma ampla gama de **recursos institucionais** para a realização de pesquisas clínicas. Esses recursos incluem pessoal especializado, serviços de gerenciamento das finanças, consultorias, bem como unidades de pesquisa clínica com área física adequada e equipe técnica experiente. Muitas universidades também dispõem de laboratórios centrais onde podem ser feitas as medições especializadas, os espaços e os equipamentos centralizados para armazenar materiais biológicos e imagens, os serviços de gerenciamento de bancos de dados, os centros profissionais de recrutamento, a experiência em lidar com aspectos legais e bibliotecas de formulários e documentos relacionados à pesquisa. Essa infraestrutura pode não ficar muito visível em uma instituição grande e muito espalhada, e os investigadores devem procurar saber sobre os recursos locais antes de tentarem fazer tudo por conta própria.

Iniciando o estudo

No início do estudo, o investigador principal deve finalizar o orçamento, desenvolver e assinar quaisquer contratos envolvidos, definir as funções de cada membro da equipe, contratar e treinar a equipe, obter aprovação pelo CEP, redigir o manual de operações, desenvolver e testar os formulários e questionários, desenvolver e testar a base de dados e

planejar o recrutamento dos participantes. Esse período de **início do estudo** (*study start-up*), que ocorre antes do arrolamento do primeiro participante, requer um esforço concentrado. Tempo e planejamento adequados para o início do estudo são fundamentais para conduzir uma pesquisa de alta qualidade.

O financiamento adequado para conduzir o estudo é crucial. O **orçamento** já deve estar pronto no momento em que a proposta é submetida para financiamento, muito antes do início do estudo (Capítulo 19). A maioria das universidades e centros médicos dispõem de pessoal com experiência na elaboração de orçamentos (**administrador pré-financiamento**). É aconselhável conhecer bem essa pessoa e respeitar o seu nível de estresse em relação aos prazos, procurando respeitar o cronograma estabelecido. Recomenda-se também compreender a fundo as normas relacionadas às várias fontes de financiamento.

Em geral, para gastar recursos do NIH e de outras fontes públicas de financiamento, as regras são muito mais restritivas do que para verbas da indústria ou de fundações. O valor total do orçamento em geral não poderá ser aumentado caso se descubra que o trabalho terá custo maior do que o previsto. Além disso, para transferir recursos entre diferentes rubricas de gastos (p. ex., pessoal, equipamentos, materiais de consumo, viagens) ou para reduzir substancialmente o trabalho de pessoas-chave no estudo, geralmente é necessária aprovação da agência de financiamento. As universidades e os centros médicos costumam ter funcionários cuja responsabilidade é assegurar que os fundos disponibilizados para um investigador por meio de auxílios e de contratos sejam gastos de forma correta. Esse **administrador pós-financiamento** deve preparar periodicamente relatórios e projeções que permitam ao investigador realizar ajustes no orçamento para fazer o melhor uso possível dos recursos disponíveis durante a condução do estudo, assegurando que o orçamento não será extrapolado no final do estudo. Sobrar parte da verba do financiamento ao final do estudo pode ser positivo, pois as agências de financiamento muitas vezes aprovam **extensões sem custos** que permitem o uso dessa sobra após o término formal do estudo para completar ou estender o trabalho descrito no escopo da proposta.

O orçamento de um estudo financiado por uma indústria farmacêutica é parte de um contrato que incorpora o protocolo da pesquisa e uma

definição clara das tarefas a serem executadas pelo investigador e pela indústria financiadora. **Contratos** são documentos legais que obrigam o investigador a realizar determinadas tarefas e descrevem o cronograma e o valor a ser pago para **metas específicas alcançadas**, como atingir o recrutamento esperado e submeter os relatórios sobre o andamento do estudo. Advogados da universidade ou do centro médico são necessários para elaborar esses contratos e para assegurar que eles protejam os direitos de propriedade intelectual do investigador, o acesso aos dados, os direitos de publicação e assim por diante. No entanto, os advogados geralmente não estão familiarizados com as tarefas necessárias para conduzir um determinado estudo, sendo necessária a participação do investigador, especialmente no que tange ao escopo da pesquisa e às metas a serem alcançadas.

Aprovação pelo Comitê de Ética em Pesquisa da instituição

O **CEP** deve aprovar o protocolo do estudo, o formulário de consentimento informado e os materiais de recrutamento antes que se comece a recrutar os participantes (Capítulo 14). Os investigadores devem estar familiarizados com as exigências de seu CEP local e com os prazos para obtenção da aprovação. Os membros do CEP são geralmente bastante prestativos nesses assuntos e devem ser contatados já no início para discutir questões procedimentais e decisões de delineamento que possam afetar os participantes do estudo.

Desenvolvendo o manual de operações e os formulários

O protocolo do estudo é muitas vezes expandido, dando origem a um **manual de operações** que inclui o protocolo, informações sobre a organização do estudo e suas regras e uma versão detalhada da seção do protocolo referente à metodologia (Apêndice 17A). Esse manual especifica exatamente como recrutar e arrolar os participantes e descreve todas as atividades que ocorrem em cada visita –randomização, cegamento, aferição de cada variável, procedimentos de controle de qualidade, práticas de gerenciamento de dados, plano de análise estatística e plano para o monitoramento dos dados e da segurança (Capítulo 11). Também deve incluir todos os questionários e formulários que serão usados no estudo, com instruções sobre como contatar os participantes,

realizar as entrevistas, preencher e codificar os formulários, entrar e editar os dados e coletar e processar as amostras biológicas. O manual de operações é essencial para pesquisas realizadas por uma equipe grande de pesquisadores, especialmente quando houver colaboração entre muitos investigadores em diferentes locais. Mesmo quando um único investigador faz todo o trabalho por conta própria, definições operacionais por escrito ajudam a reduzir a variação aleatória e mudanças na técnica de aferição ao longo do tempo.

O modo como os **formulários de coleta de dados** são elaborados tem forte influência sobre a qualidade dos dados e o sucesso do estudo (Capítulo 16). Antes de o primeiro participante ser recrutado, os formulários devem ser pré-testados. Quando uma entrada em um formulário envolver um julgamento por parte do respondente, definições operacionais explícitas devem ser resumidas brevemente no próprio formulário e descritas em mais detalhes no manual de operações. Os itens devem ser coerentes, e sua sequência, claramente formatada, com padrões para os pulos (veja Apêndice 15). Os pré-testes ajudam a assegurar clareza de significado e facilidade de uso. Escrever em cada página a data, o nome e o número de identificação do sujeito e do funcionário responsável pela coleta ajuda a proteger a integridade dos dados. Devem-se pré-testar no início do estudo também os formulários para coleta de dados na internet, em computadores de mão ou em outros dispositivos. Orientações para o uso desses equipamentos devem ser incluídas no manual de operações.

Delineamento do banco de dados

Antes de o primeiro participante ser recrutado, o banco de dados que será usado para inserir, editar, armazenar, monitorar e analisar os dados deve ser criado e testado. Dependendo do tipo de banco de dados e do escopo do estudo, poderão ser necessárias semanas a meses para desenvolver e testar o sistema de entrada e gerenciamento após a equipe com as habilidades necessárias ter sido identificada, contratada e treinada. Muitos centros médicos acadêmicos oferecem serviços para auxiliar os investigadores a desenvolver um banco de dados apropriado e oferecem opções de *softwares* para gerenciar o banco de dados. Estudos de grande porte muitas vezes usam serviços profissionais de delineamento e de

gerenciamento de bancos de dados, mas recomenda-se buscar aconselhamento sobre essas opções consultando especialistas da própria instituição e investigadores seniores.

Mesmo para estudos de pequeno porte, o tempo gasto no início do estudo para planejar e construir o banco de dados é um tempo bem gasto (Capítulo 16). Os investigadores que ficam ansiosos para iniciar um estudo e começam a registrar logo os dados muitas vezes fazem esse registro apenas em papel ou em *softwares* de planilhas eletrônicas, como o Microsoft Excel, em vez de utilizar um programa para gerenciamento de bancos de dados. Essa abordagem, embora seja inicialmente mais fácil, muitas vezes acaba, mais tarde, levando o investigador a despender muito mais tempo e esforços para analisar os dados. A vantagem de construir um banco de dados em uma etapa inicial do estudo é permitir ao investigador considerar desde o início quais valores são aceitáveis para cada variável e impedir ou gerar alertas para valores fora dessa faixa, ilógicos ou faltantes. Bons sistemas de entrada e gerenciamento de dados melhoram o controle de qualidade na coleta e entrada de dados e reduzem o tempo que posteriormente será necessário para sua limpeza. Entretanto, a maior vantagem desses sistemas é evitar descobrir tardiamente que um número grande de valores faltantes, fora da faixa ou ilógicos não poderá ser corrigido.

Recrutamento

Abordagens para recrutar com sucesso o número desejado de participantes do estudo são descritas no Capítulo 3. Aqui gostaríamos de enfatizar que o recrutamento em tempo apropriado é o aspecto mais difícil de muitos estudos. Tempo, equipe, recursos e especialistas apropriados são essenciais e devem ser planejados muito antes do início do estudo.

FINALIZANDO O PROTOCOLO

Pré-testes e ensaios gerais

Pré-testes e estudos-piloto são delineados para avaliar factibilidade, eficiência e custo de metodologias de estudo; reprodutibilidade e acurácia de aferições; taxas prováveis de recrutamento; e, às vezes, estimativas de desfecho e magnitudes de efeito ou associação. A natureza e a intensidade

dos pré-testes e estudos-piloto dependem do delineamento e das necessidades do estudo. Para a maioria dos estudos, uma série de pré-testes ou um pequeno estudo-piloto são mais do que suficientes, mas, para estudos de grande porte e custo elevado, um estudo-piloto maior pode ser mais adequado. Pode ser desejável gastar até 10% do orçamento do estudo para assegurar-se de que as estratégias de recrutamento irão funcionar, que as medições serão apropriadas e que as estimativas de tamanho de amostra sejam realistas.

Pré-testes são avaliações de determinados questionários, medidas ou procedimentos que a equipe realiza para testar sua funcionalidade, adequabilidade e factibilidade. Por exemplo, o pré-teste da entrada de dados e do gerenciamento do banco de dados é geralmente feito solicitando aos membros da equipe que preencham questionários deixando itens em branco, usando respostas fora da faixa de valores permitidos ou ilógicas, para então entrar esses dados e testar se o sistema de edição dos dados identifica os erros.

Antes de dar início ao estudo, é importante testar os planos para as visitas clínicas e outros procedimentos por meio de um **ensaio geral** em grande escala. O objetivo aqui é solucionar problemas referentes ao conjunto final de instrumentos e procedimentos. O que no papel aparenta ser um protocolo fácil e livre de problemas pode, quando implementado, revelar problemas práticos e logísticos, e assim o ensaio geral poderá levar a melhoras importantes na abordagem. O próprio investigador pode fazer o **papel do sujeito**, vivenciando o estudo e testando a equipe do ponto de vista do sujeito.

Pequenas revisões do protocolo após o início da coleta de dados

Por mais cuidadoso que tenha sido o delineamento do estudo e a pré-testagem dos procedimentos, uma vez iniciado o estudo, é inevitável que surjam problemas. A regra é fazer o mínimo possível de alterações. Entretanto, mudanças no protocolo podem, às vezes, fortalecer o estudo.

Ao decidir se uma **pequena alteração** melhora a integridade do estudo, geralmente são pesados os benefícios resultantes de uma melhor metodologia contra as desvantagens em alterar a uniformidade dos métodos do estudo, despender tempo e dinheiro na mudança do sistema e criar confusão para alguns membros da equipe. Decisões que envolvem

simplesmente a elaboração de uma **definição operacional** mais específica são relativamente fáceis. Por exemplo, é possível uma pessoa em abstinência do álcool há muitos anos ser incluído em um estudo que exclui pessoas com uso abusivo do álcool? Essa decisão deve ser tomada junto com os coinvestigadores, e divulgada pelos canais de comunicação adequados, como memorandos e reformulações no manual de operações, visando a garantir sua aplicação uniforme em todos os centros durante o restante do estudo. Muitas vezes, pequenos ajustes dessa natureza não requerem aprovação pelo CEP, especialmente se não envolverem mudanças no protocolo já aprovado pelo CEP; em caso de dúvidas, o investigador principal deve buscar esclarecimentos com algum membro do CEP. Quaisquer mudanças no protocolo, no formulário de consentimento informado, no manual de operações ou em outros documentos do estudo devem ser identificadas atribuindo-se ao documento revisado um novo número de versão. Além disso, devem ser tomadas medidas para assegurar que a versão em uso seja sempre a mais atual.

Revisões substanciais do protocolo após o início da coleta de dados

Alterações substanciais no protocolo de estudo, como incluir tipos diferentes de sujeitos ou mudar a intervenção ou o desfecho, podem ser um problema bastante sério. Embora possa haver bons motivos para realizar essas mudanças, elas deverão ser efetuadas com a noção de que os dados deverão ser analisados e relatados separadamente se isso levar a uma interpretação mais apropriada dos achados. As decisões envolvidas podem ser ilustradas com dois exemplos do estudo Raloxifene Use for the Heart (RUTH), um ensaio clínico multicêntrico sobre o efeito do tratamento com raloxifeno sobre eventos coronarianos em 10.101 mulheres com risco elevado de desenvolver eventos de doença coronariana. A definição inicial de desfecho primário era a ocorrência de infarto do miocárdio não fatal ou morte por doença coronariana. No início do estudo, percebeu-se que a taxa desse desfecho era menor do que a esperada, provavelmente porque novas cointervenções clínicas, como trombólise e angioplastia percutânea, diminuíram o risco de infarto. Após uma avaliação criteriosa, o Comitê Executivo do RUTH decidiu mudar o

desfecho primário para incluir também outras síndromes coronarianas agudas além do infarto do miocárdio. Essa mudança ocorreu no início do estudo; havia informação suficiente sobre potenciais eventos cardíacos para determinar se eles atendiam aos novos critérios para síndrome coronariana aguda, permitindo que fossem feitas buscas de eventos assim definidos de síndrome coronariana aguda, na base de dados do estudo, ocorridos antes da mudança na definição do desfecho (1).

Também no início do ensaio clínico RUTH, os resultados que haviam surgido do ensaio clínico Multiple Outcomes of Raloxifene Evaluation (MORE) mostraram que o raloxifeno diminuía de forma importante o risco relativo de desenvolver câncer de mama (2). Esses resultados não foram conclusivos, pois o número de mulheres que desenvolveram câncer de mama era pequeno, e havia preocupações quanto à capacidade de generalização, pois todas as mulheres no estudo MORE tinham osteoporose. Para avaliar se o raloxifeno também reduzia o risco de câncer de mama em outra população – mulheres idosas sem osteoporose –, o Comitê Executivo do estudo RUTH decidiu acrescentar o câncer de mama como um segundo desfecho principal (1).

Cada uma dessas mudanças foi substancial, exigindo alterações no protocolo, aprovação pelo CEP em cada centro do estudo, aprovação pelo FDA e revisão de um grande número de formulários e documentos do estudo. Esses são exemplos de revisões substanciais que melhoraram a factibilidade ou o conteúdo informativo do estudo sem comprometer sua integridade. Mudanças no protocolo nem sempre obtêm o mesmo sucesso que esses exemplos. Revisões substanciais só devem ser realizadas após pesagem dos prós e dos contras com os demais membros da equipe de pesquisa e com conselheiros apropriados, como o Comitê de Monitoramento de Dados e Segurança dos Participantes e a agência financiadora. O investigador deverá, então, lidar com o potencial impacto da mudança na análise dos dados e na formulação das conclusões do estudo.

Encerramento

Em algum momento, em todos os estudos longitudinais e ensaios clínicos, encerra-se o seguimento dos participantes. O período durante o qual os participantes realizam sua última visita no estudo é denominado

encerramento. O encerramento de estudos clínicos apresenta várias questões que merecem planejamento cuidadoso (3). No mínimo, na visita de encerramento, os membros da equipe devem agradecer aos participantes pelo tempo e esforço despendidos e informá-los de que sua participação foi fundamental para o sucesso do estudo. Além disso, o encerramento pode incluir as seguintes atividades:

- Os participantes (e seus médicos) devem geralmente ser informados dos resultados dos exames laboratoriais clinicamente relevantes ou de outras medidas que foram realizadas durante o estudo, seja de forma presencial na última visita (com cópia por escrito) ou posteriormente por e-mail.
- Em um ensaio clínico cego, os participantes podem ser informados sobre qual intervenção receberam, seja na última visita ou por *e-mail*, após todos os participantes terem completado sua participação e as análises principais já terem sido concluídas ou o manuscrito principal sobre os resultados já ter sido publicado.
- Uma cópia do manuscrito principal com base nos resultados do estudo e um *press release* ou outra descrição dos achados em linguagem leiga devem geralmente ser enviados aos participantes (e aos seus médicos) no momento da apresentação ou publicação dos resultados, juntamente com um número telefônico para os participantes que tiverem dúvidas.
- Após todos os participantes terem completado as visitas do estudo, podem ser convidados para uma recepção durante a qual o investigador principal irá agradecer por sua participação, discutir os resultados do estudo e esclarecer as dúvidas.

CONTROLE DE QUALIDADE DURANTE O ESTUDO

Boas práticas clínicas

Um aspecto crucial da pesquisa clínica é como assegurar que todos os aspectos do estudo sejam da melhor qualidade possível. Diretrizes para pesquisas de alta qualidade, denominadas GCPs, foram desenvolvidas

especificamente para ensaios clínicos que testam fármacos para que sejam aprovados pelo FDA ou por outras agências reguladoras, e são definidas como "um padrão internacional de qualidade ética e científica para delinear, conduzir e registrar e relatar os resultados de ensaios clínicos que envolvem a participação de sujeitos humanos. A adesão a esse padrão oferece a garantia pública de que os direitos, a segurança e o bem-estar dos sujeitos do estudo sejam protegidos" (4).

Esses princípios estão cada vez mais sendo aplicados a ensaios clínicos financiados por agências federais ou outras agências públicas de fomento e a outros delineamentos de pesquisa (Tabela 17.2). As exigências de GCPs são descritas em detalhes no *FDA Code of Federal Regulations Title 21* (4, 5). Os materiais da Conferência Internacional de Harmonização (6) fornecem diretrizes de controle de qualidade usadas por agências reguladoras na Europa, nos Estados Unidos e no Japão.

TABELA 17.2 Aspectos da condução de pesquisas clínicas que fazem parte das boas práticas clínicas

- O delineamento é apoiado por estudos pré-clínicos, estudos em animais ou outros dados, quando apropriado
- O estudo é conduzido de acordo com princípios de ética em pesquisa
- Um protocolo escrito é seguido cuidadosamente
- Os investigadores e aqueles que prestam atendimento clínico são treinados e qualificados
- Todos os procedimentos clínicos e laboratoriais atendem a padrões de qualidade
- Os dados são confiáveis e corretos
- Registros completos e corretos são mantidos
- Os métodos estatísticos são pré-definidos e cuidadosamente seguidos
- Os resultados são relatados de forma clara e apropriada

A melhor forma de implementar as GCPs é por meio de **procedimentos operacionais-padrão (POPs)** para todas as atividades relacionadas ao estudo. O protocolo do estudo, o manual de operações, o plano para as análises estatísticas e o plano de monitoramento dos dados e da segurança podem ser considerados POPs, mas, muitas vezes, não cobrem aspectos como treinamento e certificação do pessoal, desenvolvimento e testagem da base de dados ou manutenção, garantia de confidencialidade e cópias de segurança dos arquivos do estudo. Muitos centros médicos acadêmicos têm funcionários especializados em garantir adesão às GCPs e podem disponibilizar vários modelos para os POPs. A temática da condução ética da pesquisa é abordada no Capítulo 14. Neste capítulo nos centraremos

nos procedimentos do estudo e no gerenciamento de dados.

Controle de qualidade para procedimentos clínicos

É recomendável designar um membro da equipe de pesquisa para ser **coordenador do controle de qualidade**, responsável pela implementação das técnicas de controle de qualidade apropriadas para todos os aspectos do estudo, supervisão do treinamento e certificação da equipe e monitoramento do uso dos procedimentos de controle de qualidade durante o estudo. A meta é detectar possíveis problemas antes de eles ocorrerem, podendo, assim, preveni-los. O coordenador do controle de qualidade pode também estar preparado e atuar como pessoa de contato do estudo para auditorias do CEP, FDA (no Brasil, ANVISA) e da agência financiadora. O controle de qualidade inicia durante a fase de planejamento e continua no decorrer do estudo (Tabela 17.3).

TABELA 17.3 Controle de qualidade de procedimentos clínicos*

Passos que precedem o estudo	Desenvolver um manual de operações
	Definir estratégias de recrutamento
	Elaborar definições operacionais das medidas
	Elaborar instrumentos e formulários padronizados
	Elaborar sistemas de controle de qualidade
	Elaborar sistemas para cegar os participantes e os investigadores
	Designar um coordenador de controle de qualidade
	Treinar a equipe de pesquisa e documentar esse fato
	Certificar a equipe de pesquisa e documentar esse fato
Passos durante o estudo	Desempenhar liderança firme e afetiva
	Realizar reuniões de equipe regulares
	Elaborar procedimentos especiais para intervenções medicamentosas
	Recertificar a equipe de pesquisa
	Realizar revisões periódicas do desempenho
	Comparar periodicamente as medições entre diferentes técnicos e ao longo do tempo

* Os procedimentos clínicos incluem medição da pressão arterial, entrevistas estruturadas, revisão de prontuários, etc.

- **Manual de operações.** O manual de operações é um aspecto muito importante do controle de qualidade (Apêndice 17A). Para ilustrar, considere a aferição da estatura em um estudo no qual a mudança na estatura será usada como preditor da osteoporose. Uma vez que a

aferição da estatura é um desfecho parcialmente subjetivo para o qual não existe padrão-ouro factível, o manual de operações deveria fornecer instruções específicas sobre o dispositivo de aferição a ser usado (marca e modelo do estadiômetro), bem como instruções para preparar o participante para a aferição (remover os sapatos), posicioná-lo no dispositivo e realizar a aferição.

- **Calibração, treinamento e certificação.** Os dispositivos de aferição (balanças, estadiômetros, equipamentos de imagem, equipamentos laboratoriais, etc.) devem ser calibrados profissionalmente antes do início do estudo e periodicamente ao longo do estudo. O treinamento padronizado da equipe é essencial para pesquisas de alta qualidade. Toda a equipe deveria receber treinamento adequado antes do início do estudo e receber certificação sobre sua competência em relação aos procedimentos e aferições importantes. Em relação à aferição da estatura, por exemplo, os membros da equipe podem ser treinados em cada aspecto dela, exigindo-se que obtenham medidas satisfatórias em participantes que não farão parte da população de pesquisa e cuja estatura é conhecida. O procedimento de certificação deve ser suplementado durante o estudo por recertificações previamente agendadas, e um registro dos treinamentos, certificações e recertificações deve ser mantido no centro de estudo.

- **Revisão do desempenho.** Os supervisores devem revisar com uma certa periodicidade a forma como os procedimentos clínicos são realizados, acompanhando em silêncio visitas clínicas ou ligações telefônicas representativas. Após obter a permissão do participante do estudo, o supervisor pode estar presente, em silêncio, em pelo menos um exemplo completo de cada tipo de entrevista e de procedimento técnico realizado por cada membro de sua equipe. Isso pode gerar um certo desconforto no início, mas logo a equipe se habitua. É bastante útil empregar durante essas observações um *checklist* padronizado (fornecido com antecedência e baseado no protocolo e no manual de operações). Mais tarde, a comunicação entre o supervisor e o membro da equipe de pesquisa pode ser facilitada revisando-se o *checklist* e discutindo-se de forma positiva e não pejorativa quaisquer problemas

de controle de qualidade que tenham sido identificados. O momento em que essas observações são realizadas e os seus resultados devem ser anotados no registro do treinamento.

O envolvimento de **pares** da equipe de pesquisa nas revisões ajuda a construir um espírito de grupo e garante a consistência na aplicação de abordagens padronizadas entre membros da equipe que desempenham as mesmas funções. Uma vantagem de se usar colegas como observadores nesse sistema é que todos os membros da equipe se sentem responsáveis pelo controle de qualidade. Outra vantagem é que o observador aprende tanto quanto a pessoa que está sendo observada.

- **Relatórios periódicos.** É importante **tabular os dados** sobre qualidade técnica de procedimentos clínicos e métodos de aferição em intervalos regulares. Essa tabulação pode apontar para a ocorrência de medições faltantes, inacuradas ou variáveis. Diferenças entre níveis médios de pressão arterial observados por membros de uma equipe que mede a pressão arterial nos últimos dois meses, por exemplo, podem apontar diferenças nas técnicas utilizadas. Do mesmo modo, uma mudança gradual no desvio-padrão dos conjuntos de leituras ao longo de alguns meses pode indicar uma mudança na técnica de medição. Relatórios periódicos devem também abordar o sucesso do recrutamento, o respeito aos prazos para entrada de dados, a proporção de variáveis faltantes ou fora da faixa permitida, o momento para realizar as consultas (*queries*) no monitoramento dos dados e o sucesso do seguimento e da adesão à intervenção.

- **Procedimentos especiais para intervenções medicamentosas.** Ensaios clínicos sobre intervenções medicamentosas, especialmente aqueles que utilizam cegamento, exigem atenção especial no controle de qualidade da rotulagem, entrega e armazenamento; dispensação do medicamento; e recolhimento e descarte adequado dos medicamentos não usados. O fornecimento correto do medicamento e sua dosagem podem ser assegurados planejando-se criteriosamente junto ao fabricante ou farmácia da pesquisa sobre a natureza da abordagem de distribuição, supervisionando-se sua implementação e, eventualmente, testando-se a composição das medicações mascaradas para assegurar que elas

correspondam ao fármaco correto. Estudos sobre medicamentos também requerem procedimentos claros e registros do recebimento das medicações do estudo, do seu armazenamento, distribuição e retorno pelos participantes.

Controle de qualidade dos procedimentos laboratoriais

A qualidade dos procedimentos laboratoriais pode ser controlada usando-se várias das abordagens descritas na Tabela 17.3 para procedimentos clínicos. Além disso, o fato de se estar coletando amostras dos sujeitos (criando a possibilidade de erros de rotulagem) e o caráter técnico de exames laboratoriais levam a diversas estratégias:

- **Atenção à rotulagem.** Quando a amostra de sangue de um sujeito é rotulada por engano com o nome de outro indivíduo, pode ser impossível corrigir ou até mesmo descobrir mais tarde o erro. A única solução é prevenir o erro. Para evitar **rotulagem incorreta e erros de transposição**, o nome do sujeito e seu número devem ser checados minuciosamente no momento de rotular a amostra. Conjuntos de rótulos para os tubos de sangue e para prontuários podem ser impressos em computador, a fim de agilizar o processo de rotulagem e evitar erros que podem acontecer quando números são escritos à mão. Um procedimento aconselhável para a transferência de soro de um tubo para outro é rotular previamente o tubo novo, manter os dois tubos lado a lado e ler um deles em voz alta ao mesmo tempo em que se verifica o outro. A rotulagem também pode ser automatizada por meio de **códigos de barras**.

- **Cegamento.** A tarefa de cegar o observador é fácil quando se trata de medições em amostras previamente coletadas. É sempre aconselhável rotular as amostras de modo que o técnico não tenha conhecimento do grupo do sujeito ou dos valores de outras variáveis importantes. Mesmo para procedimentos aparentemente objetivos, como determinação automatizada da glicemia, essa precaução reduz imensamente as chances de vieses, além de facilitar a elaboração de uma seção convincente de métodos no relato dos resultados. No entanto, quando se cegam os técnicos do laboratório, deve haver procedimentos claros para

relatar resultados anormais a um membro da equipe que esteja qualificado para revisar os resultados e decidir se o participante deveria ser notificado ou se outra ação deveria ser tomada. Em ensaios clínicos, deve também haver estratégias para o descegamento (às vezes emergencial) se medições laboratoriais indicarem anormalidades que poderiam estar associadas à intervenção do estudo e que requerem ação imediata.

- **Duplicatas mascaradas, pools-padrão e medidas de consenso.** Quando amostras ou imagens de um estudo multicêntrico são enviadas a um laboratório central para análise química ou interpretação, pode ser conveniente que se enviem duplicatas mascaradas utilizando-se o mesmo sistema. Essas duplicatas podem ser uma segunda amostra de um subconjunto aleatório de sujeitos aos quais é dado um número de identificação fictício. Essa estratégia fornece uma estimativa da precisão das técnicas laboratoriais. Outra abordagem para amostras de soro que podem ser congeladas consiste em preparar inicialmente um *pool* de soro e enviar periodicamente por meio do sistema certas quantidades dele, mascaradamente rotuladas, com números de identificação fictícios. As medições iniciais do *pool* de soro, conduzidas com as melhores técnicas disponíveis, estabelecem seus valores; esses valores do *pool* são, então, usados como padrão-ouro durante o estudo, possibilitando estimativas de acurácia e precisão. Uma terceira abordagem, usada para medições com variabilidade inerente, como o exame de Papanicolau ou a mamografia, é envolver dois leitores independentes, cegados. Se ambos concordarem dentro de uma margem aceitável de diferença, o resultado é tido como correto. Resultados discordantes podem ser resolvidos mediante discussão ou consenso, ou buscando-se a opinião de um terceiro leitor.

- **Contratos com laboratórios comerciais.** Em alguns estudos, análises biológicas em sangue, soro, células ou tecidos são feitas mediante contrato com laboratórios comerciais. O laboratório deve dispor de licenciamento e certificação adequados e uma cópia desses documentos deve estar disponível no centro de estudo. Laboratórios comerciais devem fornecer dados sobre a reprodutibilidade de suas aferições, como

coeficientes de variação, assegurar o cumprimento de prazos e oferecer procedimentos padronizados para lidar com amostras codificadas, notificar os investigadores sobre resultados anormais e transferir os dados para o banco principal.

Controle de qualidade do gerenciamento de dados

O investigador deve montar e pré-testar o sistema de gerenciamento de dados antes de dar início ao estudo. Isso inclui elaborar os formulários para o registro das aferições; escolher os equipamentos de informática e os *softwares* para entrada, edição e gerenciamento dos dados; definir os parâmetros de edição de dados para entradas faltantes, fora de faixa ou ilógicas; testar o sistema de gerenciamento de dados; e planejar tabulações simuladas para assegurar que as variáveis apropriadas sejam coletadas (Tabela 17.4).

TABELA 17.4 Controle de qualidade do gerenciamento de dados: passos que precedem o estudo

Ser parcimonioso: coletar apenas as variáveis necessárias
Selecionar os equipamentos de informática e os softwares para gerenciamento do banco de dados apropriados
Programar o banco de dados para emitir alertas sobre valores faltantes, fora da faixa permitida ou ilógicos
Testar o banco de dados usando valores faltantes, fora da faixa permitida e ilógicos
Planejar as análises e testá-las com tabulações simuladas
Elaborar formulários em papel ou eletrônicos que sejam
Autoexplicativos
Coerentes (p. ex., opções de escolha múltipla exaustivas e mutuamente exclusivas)
Claramente formatados para entrada de dados, com setas indicando os pulos
Impressos em caixa baixa, usando caixa alta, sublinhado e negrito para ênfase
Esteticamente adequados e de fácil leitura
Pré-testados e validados (ver Capítulo 15)
Rotulados em cada página com data, nome, número de identificação e/ou código de barras

- **Dados faltantes.** Dados faltantes podem ser desastrosos caso afetem uma grande proporção das medições. Mesmo poucos valores faltantes podem, em alguns casos, enviesar as conclusões do estudo. Por exemplo, um estudo sobre sequelas a longo prazo de uma cirurgia que apresenta taxa de mortalidade tardia de 5% pode subestimar consideravelmente essa complicação se 10% dos participantes forem perdidos e se a morte for uma razão comum para essas perdas.

Conclusões errôneas devidas a dados faltantes podem, às vezes, ser corrigidas *a posteriori* –, o que, nesse caso, exigiria um grande esforço para localização dos pacientes perdidos –, mas, muitas vezes, a medição não pode ser reposta. Existem técnicas estatísticas para **imputar valores faltantes** a partir de outras informações da linha de base ou das visitas de seguimento ou a partir da média dos valores dos demais participantes. Embora essas técnicas sejam úteis, particularmente para análise multivariada em que o acúmulo de dados faltantes pode inviabilizar o uso de uma grande fração dos sujeitos, elas não garantem conclusões livres de vieses de não resposta caso o número de observações faltantes seja grande.

A única solução adequada é delinear e conduzir o estudo de uma forma que evite dados faltantes. Por exemplo, pode-se designar um membro da equipe para avaliar a completude dos formulários antes de o participante deixar a clínica, elaborar interfaces eletrônicas para entrada de dados que não permitam pular entradas e programar o banco de dados de modo a criar alertas para a equipe de pesquisa cada vez que aparecer um dado faltante (Tabela 17.5). Medições clínicas faltantes devem ser abordadas enquanto o participante ainda está na clínica, quando ainda é relativamente fácil corrigir os erros.

TABELA 17.5 Controle de qualidade do gerenciamento de dados: passos durante o estudo

Assinalar ou verificar omissões e erros importantes enquanto o participante ainda estiver na clínica. Assegurar-se de que:
Não há erros ou problemas de transposição do número de identidade, código do nome, ou data em cada página
Todos os formulários corretos para a visita especificada foram preenchidos
Não há nenhuma entrada faltante ou padrão incorreto de pulos
As entradas são legíveis
Os valores de variáveis importantes estão dentro da faixa permitida
Os valores de variáveis importantes são consistentes uns com os outros (p. ex., idade e data de nascimento
Realizar periodicamente distribuições de frequências e medidas de variância para identificar valores aberrantes
Criar outras tabulações periódicas para descobrir erros (ver Apêndice 17B)

- **Dados inacurados e imprecisos.** Este é um problema insidioso, que muitas vezes permanece sem ser descoberto, especialmente quando mais de uma pessoa está envolvida na realização das aferições. No pior cenário, o investigador planeja o estudo e deixa a coleta de dados para

ser realizada por seus assistentes de pesquisa. Ao retornar para analisar os dados, algumas das medições podem apresentar viés importante devido ao uso repetido de uma técnica inadequada. Esse problema torna-se particularmente grave quando os erros nos dados não podem ser detectados *a posteriori*. O investigador irá pressupor que as variáveis representam o que deveriam representar e, ignorando o problema, poderá chegar a conclusões equivocadas sobre o estudo.

O treinamento e a certificação da equipe, as revisões de desempenho e a avaliação continuada das médias e faixas de valores dos dados gerados por diferentes membros da equipe podem ajudar a identificar ou prevenir esses problemas. A **edição computadorizada** desempenha papel importante, usando sistemas de entrada e gerenciamento de dados programados para emitir alertas ou não permitir a submissão de formulários com dados faltantes, inconsistentes ou fora da faixa permitida. Um procedimento padronizado deve estar disponível para corrigir os dados originais em qualquer formulário de dados. Geralmente ele deveria ser aplicado tão cedo na coleta de dados quanto possível, com um processo que inclua riscar a informação original (não apagá-la), assinar e colocar a data da mudança. Processos semelhantes deveriam ser incluídos nos sistemas eletrônicos para entrada e edição de dados. Isso proporciona um **rastreamento de operações** (*audit trail*) eletrônico para justificar mudanças nos dados e prevenir fraudes.

A tabulação e inspeção periódica das distribuições de frequências de variáveis importantes em intervalos regulares permite que o investigador avalie a completude e a qualidade dos dados em um momento em que a correção de erros ocorridos no passado ainda é possível (p. ex., contatando o participante por *e-mail* ou telefone ou solicitando que ele retorne à clínica) e quando erros futuros podem ser prevenidos. Uma lista útil de tópicos para relatórios de controle de qualidade é fornecida no Apêndice 17B.

- **Dados fraudulentos.** Os investigadores clínicos que lideram uma equipe de pesquisa devem estar preparados para se deparar com indivíduos inescrupulosos que optem por fabricar informações para simplificar a conclusão de suas tarefas. A fim de evitar esses eventos desastrosos, é importante escolher cuidadosamente os colaboradores e

outros membros da equipe; estabelecer um relacionamento sólido com os membros do grupo, assegurando um código de comportamento ético explicitamente compreendido por todos; ficar alerta à possibilidade de fraude ao examinar os dados; e eventualmente verificar a fonte primária dos dados, certificando-se de sua veracidade.

Estudos colaborativos multicêntricos

Muitas questões de pesquisa requerem um número maior de sujeitos que o disponível em um único centro, exigindo estudos colaborativos com equipes de vários locais. Às vezes, esses locais concentram-se na mesma cidade ou Estado, e um único investigador poderá supervisionar todas as equipes. Entretanto, estudos colaborativos são, muitas vezes, conduzidos por investigadores que moram em cidades separadas por milhares de quilômetros, com financiamento separado e diferentes estruturas administrativas e regulatórias.

Estudos dessa natureza requerem medidas especiais para assegurar que todos os centros estejam desenvolvendo o mesmo estudo e produzindo dados comparáveis que possam ser combinados na análise dos resultados. Um **centro de coordenação** estabelece uma rede de comunicação; coordena o desenvolvimento do manual de operações, dos formulários e outros aspectos inerentes ao controle de qualidade do estudo; treina o pessoal responsável pelas medições em cada centro e supervisiona o gerenciamento, a análise e a publicação dos dados. Estudos colaborativos geralmente possuem sistemas de entrada distribuída de dados conectados por meio da internet.

É também necessário estabelecer um sistema de governança com um **comitê diretivo** constituído pelos investigadores principais e representantes da instituição financiadora e com vários subcomitês. Um desses **subcomitês** deve ser responsável pelos assuntos referentes ao **controle de qualidade**, desenvolvimento de procedimentos de padronização e sistemas para treinamento, certificação e revisão de desempenho. Essas tarefas tendem a ser complicadas e onerosas, exigindo **treinamento centralizado** para os membros relevantes de cada centro, **visitas aos centros** para revisão do desempenho e auditorias dos dados por membros do centro de coordenação e outros colaboradores (Apêndice 17B). Os demais subcomitês geralmente incluem grupos que

supervisionam o **recrutamento** e as **atividades clínicas**, um grupo que revisa e aprova as **publicações e apresentações** e um que avalia propostas para estudos suplementares.

Em um estudo multicêntrico, mudanças nas definições operacionais e em outros métodos do estudo muitas vezes resultam de questionamentos levantados por um centro clínico que são respondidos pelo pessoal ou comitê apropriado do estudo e postados na página do estudo na internet em uma lista de modo a assegurar que todos os envolvidos no estudo estejam cientes das mudanças. Se houver um número significativo de mudanças, páginas revisadas e datadas no manual de operações e nos demais documentos do estudo devem ser preparadas incluindo essas mudanças. Estudos pequenos, com um único centro, podem seguir uma abordagem semelhante, fazendo observações sobre mudanças que são datadas e incluídas no manual de operações.

Consideração final

Um erro bastante comum em pesquisa é coletar uma **quantidade excessiva de dados**. Como a única oportunidade para medir as variáveis basais é durante a linha de base, o investigador fica tentado a incluir todos os dados que possivelmente serão de interesse, fazendo com que haja uma tendência de se fazer mais visitas de seguimento e de se coletar mais dados nessas visitas do que serão úteis. Assim, os investigadores tendem a coletar uma quantidade de dados muito maior do que aquela que será analisada ou publicada.

Um problema com essa abordagem é o tempo e os custos associados à medição de variáveis de menor interesse; os sujeitos se sentem entediados e cansados e o controle de qualidade das medições mais importantes se deteriora. Outro problema é o tamanho e a complexidade da base de dados que dificulta o controle de qualidade e a análise. É importante questionar a necessidade de cada variável que será coletada e eliminar muitas daquelas opcionais. Redundâncias intencionais podem incrementar a validade das variáveis importantes, mas parcimônia é a regra.

RESUMO

1. A implementação bem-sucedida do estudo inicia pela **montagem dos**

recursos, incluindo **área física**, **equipe** e **financiamento** para **dar início ao estudo**, o que exige forte liderança por parte do IP.

2. No **início do estudo,** é necessário cuidar do **orçamento**, obter **aprovação do CEP** e finalizar o **protocolo** e o **manual de operações** mediante um processo de **pré-testes** para avaliar a adequação e factibilidade dos planos de **recrutamento**, **intervenções, aferição das variáveis preditoras e de desfecho, formulários e banco de dados**; a meta é minimizar a necessidade de revisões no protocolo após o início da coleta de dados.
3. **Revisões pequenas no protocolo** após o início do estudo, como acrescentar um item a um questionário ou modificar uma definição operacional, são **fáceis de implementar**, embora às vezes possa ser necessária aprovação pelo CEP e possa haver impacto sobre a análise dos dados.
4. **Revisões maiores no protocolo** após o início do estudo, como mudar a natureza da intervenção, os critérios de inclusão ou o desfecho principal, podem ter **implicações importantes** e devem ser realizadas com cuidado e com a aprovação de órgãos importantes, como o Comitê de Monitoramento de Dados e Segurança dos Participantes, o CEP e a instituição financiadora.
5. São necessários procedimentos de **encerramento**, para informar adequadamente aos participantes do estudo sobre os achados e para coordenar a transição do cuidado e suas implicações.
6. O **controle de qualidade** durante o estudo deve ser assegurado por meio de uma abordagem sistemática, sob a supervisão de um **coordenador de controle de qualidade**, seguindo os princípios das **GCPs** e incluindo:
   a. POPs, com um **manual de operações**; **treinamento e certificação da equipe**, **revisão do desempenho**, **relatórios periódicos** (sobre recrutamento, adesão às visitas e medições) e **encontros regulares da equipe**.
   b. Controle de qualidade para **procedimentos laboratoriais** – **cegamento** e **rotulagem** sistemática das amostras dos participantes do estudo e uso de *pools*-**padrão**, **duplicatas mascaradas** e **medidas de consenso**.
   c. Controle de qualidade do **gerenciamento de dados** – elaborar

formulários e sistemas eletrônicos para permitir controle da **completude**, **acurácia** e **integridade** da coleta, entrada, edição e análise dos dados.

7. **Estudos colaborativos multicêntricos** possuem **subcomitês** e outros sistemas descentralizados para gerenciar o estudo e o controle de qualidade.

APÊNDICE 17A

Exemplo de sumário de um manual de operações[1]

Capítulo 1. Protocolo do estudo
Capítulo 2. Organização e políticas
  Unidades participantes (centros clínicos, laboratórios, centro de coordenação, etc.), investigadores e equipe
  Administração e governança (comitês, agência financiadora, monitoramento de dados e segurança, etc.)
  Diretrizes sobre políticas (publicações e apresentações, estudos suplementares, conflitos de interesse, etc.)
Capítulo 3. Recrutamento
  Critérios de elegibilidade e de exclusão
  Plano de amostragem
  Abordagens de recrutamento (publicidade, contatos para referência, rastreamento, etc.)
  Consentimento informado
Capítulo 4. Visitas à clínica
  Conteúdo da visita da linha de base
  Conteúdo e periodicidade das consultas de seguimento
  Procedimentos de seguimento para não respondentes
Capítulo 5. Procedimentos de randomização e cegamento
Capítulo 6. Variáveis preditoras
  Procedimentos de medição
  Intervenção, incluindo procedimentos para rotulagem, dispensação e manuseio das medicações
  Avaliação da adesão
Capítulo 7. Variáveis de desfecho
  Avaliação e adjudicação dos desfechos principais
  Avaliação e gerenciamento de outros desfechos e eventos adversos
Capítulo 8. Controle de qualidade
  Quadro geral e responsabilidades
  Treinamento dos procedimentos

- Certificação da equipe
- Manutenção do equipamento
- Revisão por pares e visitas aos centros
- Relatórios periódicos

Capítulo 9. Gerenciamento de dados

- Coleta e registro dos dados
- Entrada dos dados
- Edição, armazenamento e cópia de segurança
- Confidencialidade

Capítulo 10. Planos para análise de dados

Capítulo 11. Diretrizes para Monitoramento dos Dados e da Segurança

Apêndices

- Cartas aos sujeitos, aos médicos e assim por diante
- Questionários, formulários
- Detalhamento de procedimentos, critérios e assim por diante
- Materiais para o recrutamento (anúncios, materiais informativos, cartas, etc.)

APÊNDICE 17B
Tabelas e *checklists* para controle de qualidade

I. Tabulações para monitoramento de características de desempenho[2]

A. Características dos centros clínicos

1. Recrutamento
   a. Número de participantes identificados para arrolamento; número excluído e tabulação das razões para exclusão
   b. Gráfico cumulativo do número de recrutados comparado ao número necessário para atingir o objetivo do recrutamento
2. Seguimento
   a. Número de exames de seguimento completados para cada retorno previsto; número de sujeitos examinados em um determinado período
   b. Medidas da adesão à intervenção do estudo, às visitas e às aferições
   c. Número de abandonos e de participantes não localizados no seguimento
3. Quantidade e qualidade dos dados
   a. Número de formulários completados, número de formulários que gerou mensagens de edição e número de consultas de edição (*edit queries*) não atendidas, tempo até a conclusão das consultas
   b. Número de formulários faltantes, número ou proporção de variáveis faltantes
4. Adesão ao protocolo
   a. Número de sujeitos inelegíveis arrolados
   b. Resumo dos dados das contagens de comprimidos e de outros testes de adesão para cada grupo de tratamento

B. Características do centro de dados

1. Número de formulários recebidos e número aguardando entrada
2. Lista cumulativa das alterações de codificação e do protocolo
3. Cronograma indicando as tarefas cumpridas e as não concluídas

C. Características do laboratório central

1. Número de amostras recebidas e número de amostras analisadas

2. Número de amostras identificadas inadequadamente, perdidas ou destruídas
3. Número de amostras que requerem reanálise com tabulação das razões
4. Média e variância das diferenças entre duplicatas mascaradas ao longo do tempo e análises das tendências temporais baseadas em determinações repetidas de padrões de valor conhecido

D. Características dos centros de leitura

1. Número de registros recebidos e lidos
2. Número de registros recebidos apresentando rotulagem inadequada ou outras deficiências (tabular deficiências)
3. Análises de leituras repetidas para verificação da reprodutibilidade e como meio de monitoramento de tendências temporais no processo de leitura

II. Componentes das visitas aos centros:

A. Visita ao centro clínico

1. Reunião fechada entre os visitantes e o investigador principal
2. Reunião entre os visitantes e a equipe do centro clínico
3. Inspeção das dependências para os exames e para armazenamento dos registros
4. Comparação entre o conteúdo de formulários escolhidos aleatoriamente e os pertencentes ao arquivo de dados do computador
5. Revisão do arquivo dos formulários de dados e registros relacionados para avaliar a completude e a segurança contra perda ou uso inadequado
6. Observação da equipe clínica responsável por procedimentos específicos
7. Verificação dos manuais de operações, formulários e outros documentos arquivados na clínica para avaliar se estão atualizados
8. Observação ou discussão para revisão de certos procedimentos (p. ex., a série de exames necessários para determinar a elegibilidade dos participantes)
9. Conversas com participantes reais do estudo durante ou após o arrolamento para verificar o processo de consentimento informado
10. Conversas fechadas com o pessoal de apoio, a fim de avaliar suas

práticas e filosofia de trabalho em relação à coleta de dados

11. Reunião fechada com o investigador principal para discussão dos problemas identificados

B. Visita ao centro de dados

1. Revisão dos métodos para fazer o inventário dos dados recebidos dos centros clínicos
2. Revisão dos métodos para gerenciamento e verificação de dados
3. Avaliação da adequação dos métodos utilizados para o arquivamento e armazenamento de registros em papel oriundos dos centros clínicos, incluindo segurança do local e métodos de proteção contra perda ou uso indevido dos documentos
4. Revisão dos recursos de informática disponíveis
5. Revisão dos métodos de randomização e das salvaguardas contra falhas no processo de randomização
6. Revisão dos procedimentos de edição de dados e rastreamento de operações
7. Revisão da estrutura dos arquivos eletrônicos de dados e métodos para manutenção do banco de dados para análise
8. Revisão dos métodos de programação tanto para o gerenciamento quanto para a análise dos dados, incluindo uma avaliação da documentação da programação
9. Comparação entre as informações contidas nos formulários originais do estudo e aquelas nos arquivos eletrônicos de dados
10. Revisão dos métodos de geração de arquivos de dados para análise e relatórios de dados relacionados
11. Revisão dos métodos para criação de cópias de segurança do arquivo de dados principal
12. Revisão do arquivo-mestre dos documentos-chave do estudo – manuais, formulários de dados, atas dos comitês do estudo, etc., objetivando a completude

REFERÊNCIAS

1. Mosca L, Barrett-Connor E, Wenger NK, et al. Design and methods of the Raloxifene Use for The Heart (RUTH) Study. *Am J Cardiol* 2001;88:392–395.
2. MORE Investigators. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple outcomes of raloxifene evaluation. *JAMA* 1999;281:2189–2197.

3. Shepherd R, Macer JL, Grady D. Planning for closeout–From day one. *Contemp Clin Trials* 2008; 29:136–139
4. http://www.fda.gov/downloads/Drugs/Guidances/ucm073122.pdf
5. FDA Regulations Relating to Good Clinical Practice and Clinical Trials. Available at: www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/ucm114928.htm
6. Information about Good Clinical Practices in the European Medicines Agency International Conference on Harmonization. Available at: http://www.ich.org or at http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000035.jsp&murl=menus/regulations

---

1 Este é um modelo para um estudo multicêntrico de grande porte. O manual de operações para um estudo menor pode ser menos elaborado.

2 As tabelas devem conter resultados para todo o período de estudo, e, quando apropriado, para o período compreendido desde a produção do último relatório. Taxas e comparações entre membros da equipe e unidades participantes devem ser fornecidas quando apropriado.

CAPÍTULO 18

Estudos comunitários e internacionais

Norman Hearst e Thomas E. Novotny

As pesquisas clínicas são, em sua maioria, conduzidas em centros médicos universitários ou em outras instituições médicas de grande porte. Esses lugares oferecem diversas vantagens ao desenvolvimento da pesquisa, dentre as quais a mais óbvia é a presença de pesquisadores experientes. Aspectos como cultura acadêmica estabelecida, reputação e infraestrutura de pesquisa facilitam o trabalho de todos, desde o investigador iniciante até o professor titular. Sucesso atrai mais sucesso, e, por essa razão, a pesquisa clínica se concentra em centros de excelência. Este capítulo, no entanto, trata das pesquisas conduzidas fora desses centros.

Definimos **pesquisa comunitária** como aquela conduzida fora dos centros médicos usuais e destinada a atender às necessidades da comunidade onde ela será desenvolvida. A **pesquisa internacional**, principalmente nos países pobres, pode envolver muitos dos mesmos desafios de responder às necessidades locais e de estabelecer uma infraestrutura de pesquisa em um local em que ela antes não existia. Para realizar estudos internacionais, é também necessário compreender as inúmeras complexidades políticas, burocráticas e culturais que podem surgir nesse contexto. Estudos comunitários e internacionais muitas vezes envolvem **colaboração** entre investigadores locais e colegas de um centro de pesquisa já estabelecido. Colaborações como essa são fundamentais para resolver problemas de saúde antigos ou emergentes, em nível global ou local, e podem constituir oportunidades extraordinárias para crescimento pessoal e aprendizado mútuo. Entretanto, podem ser **desafiadoras** devido à **distância física** entre os investigadores, às

diferenças culturais envolvendo os participantes, às **questões políticas** envolvendo instituições locais e nacionais e às **limitações de financiamento**, tanto no nível da instituição que o oferece quanto daquela que o recebe.

POR QUE PESQUISA COMUNITÁRIA E INTERNACIONAL?

A pesquisa colaborativa é, às vezes, a única forma de abordar questões sobre cenários especiais, doenças novas ou reemergentes ou determinadas populações. Pesquisas em centros médicos acadêmicos frequentemente focam em prioridades clínicas ou de ciência básica muito distantes das necessidades da comunidade local, distanciando-se ainda mais dos problemas de saúde globais que afetam grandes segmentos da população mundial. Para esses problemas globais, são necessários esforços coletivos, seja por motivos humanitários ou porque as fronteiras nacionais, estaduais ou locais não isolam as comunidades dos efeitos de tais problemas. O "**desequilíbrio 10/90**" na pesquisa em saúde, situação na qual 90% da carga global de doença recebe apenas 10% do investimento global em pesquisa (1), é uma forte justificativa para mais estudos colaborativos sobre problemas importantes de saúde de **países de baixa ou média renda**. É, portanto, necessário reforçar a **capacidade** de pesquisa em países e comunidades de baixa e média renda, por meio de colaboração internacional e nacional. Isso inclui atenção cuidadosa para o desenvolvimento de processos de revisão institucionais e para a proteção dos sujeitos humanos.

Participar na pesquisa traz benefícios à comunidade e aos pesquisadores que vão além da informação coletada em um estudo. Relações duradouras, senso de realização pessoal e, talvez, crescimento econômico podem resultar de pesquisas comunitárias feitas com cuidado e preocupação com o bem público.

Questões locais

Muitas questões de pesquisa exigem respostas que só podem ser obtidas por meio de **pesquisa local**. Dados nacionais ou estaduais podem não representar adequadamente a carga de doenças ou a distribuição de fatores de risco em uma comunidade. Intervenções, especialmente as que buscam

alterar comportamentos, podem ter efeitos diferentes em diferentes cenários. Por exemplo, a efetividade em termos de saúde pública da promoção do uso do preservativo como estratégia de prevenção de HIV/AIDS é muito diferente nos Estados Unidos em comparação com a África (2). Para abordagens que se enquadrem às necessidades de um local, são necessários métodos adequados a pesquisas locais (Tabela 18.1).

TABELA 18.1 Exemplos de questões de pesquisa que requerem pesquisas locais

Quais são as taxas de uso de assento infantil e de cinto de segurança em um bairro de baixa renda em Chicago?
Quais são os padrões de resistência a tuberculostáticos em amostras de escarro colhidas em Uganda?
Qual é o impacto de campanhas no ambiente de trabalho para prevenção de doenças sexualmente transmissíveis em trabalhadores agrícolas itinerantes no Texas?
Qual é a proporção dos casos de doença arterial coronariana associada ao tabagismo em mulheres no Brasil?

Dados biológicos sobre a fisiopatologia de uma doença e sobre a efetividade de um tratamento são normalmente generalizáveis a uma grande variedade de populações e culturas. Entretanto, poderão ocorrer diferenças raciais, culturais ou genéticas ou diferenças baseadas na etiologia regional da doença, que requerem pesquisas locais. Por exemplo, a eficácia de medicamentos anti-hipertensivos difere entre pessoas com ascendência africana e europeia (3), os agentes causadores da pneumonia e seus padrões de sensibilidade aos antimicrobianos na Bolívia são diferentes dos de Boston e a percepção de saúde, cuidados de saúde e doença podem diferir enormemente em diferentes comunidades (4).

Maior capacidade de generalização

Uma vantagem da **pesquisa comunitária** é sua capacidade de produzir resultados mais **generalizáveis**. Por exemplo, pacientes com dores nas costas atendidos em hospitais de referência são bastante diferentes daqueles pacientes também com dores nas costas atendidos em centros de atenção primária. Estudos sobre a história natural ou a resposta terapêutica de dor nas costas realizados em centros de atenção terciária podem ter aplicação limitada na prática clínica comunitária.

Em parte em resposta a esse problema, foram desenvolvidas várias **redes de pesquisa baseadas em serviços clínicos**, a partir de parcerias de

pesquisa para o estudo de questões de interesse mútuo (5). Um exemplo é um estudo sobre a resposta ao tratamento de pacientes portadores da síndrome do túnel do carpo atendidos em serviços de atenção primária (6). A maioria dos pacientes demonstrou melhora com tratamento conservador e somente alguns necessitaram de encaminhamento a um especialista ou exames diagnósticos sofisticados. Estudos anteriores haviam recomendado intervenção cirúrgica precoce para a síndrome do túnel do carpo, com base em estudos de pacientes tratados em um grande centro de referência.

As questões sobre capacidade de generalização são também importantes na **pesquisa internacional**. Os resultados de pesquisas em um país nem sempre se aplicam em outro país. Por outro lado, embora os resultados normalmente sejam mais bem aplicados nos locais onde a pesquisa foi realizada, eles podem também ser relevantes para populações de imigrantes que se originaram daquele país. Tais populações de pessoas deslocadas vêm ganhando importância crescente em um mundo globalizado que teve 214 milhões de imigrantes internacionais no ano de 2010 (7). A globalização trouxe a necessidade de uma perspectiva mais ampla sobre o risco de doenças e sobre estratégias para pesquisas colaborativas que buscam abordar doenças que cruzam fronteiras com enorme facilidade.

Construindo capacidade local

A pesquisa clínica não deve ser propriedade exclusiva de alguns centros médicos acadêmicos. As prioridades dos investigadores, nesses locais, refletem as prioridades das instituições financiadoras, as questões que eles encontram na sua prática clínica diária e o que acreditam ser de importância científica ou econômica em geral. Por isso, ao se conduzir uma pesquisa em cenários comunitários e internacionais, garante-se que questões de importância local deverão ser priorizadas (8).

O valor da **participação da comunidade** em uma pesquisa vai além das informações específicas coletadas no estudo. Conduzir uma pesquisa tem um efeito cascata positivo à medida que incrementa os padrões acadêmicos locais e incentiva o pensamento criativo e independente. Cada projeto promove habilidades e confiança, fazendo com que os pesquisadores da comunidade se sintam participantes no processo

científico e não meros consumidores do conhecimento produzido em outro local. Isso, por sua vez, estimula a produção de mais pesquisa. Além disso, a participação em uma pesquisa pode trazer recursos intelectuais e financeiros à comunidade, incentivando a autossuficiência e fortalecendo o poder local.

PESQUISA COMUNITÁRIA

Em termos teóricos, o processo de dar início a uma pesquisa comunitária é o mesmo que para outros tipos de pesquisa. A abordagem geral descrita neste livro se aplica tanto para uma pequena cidade rural norte-americana ou do Nepal quanto para São Francisco ou Londres. Na prática, o grande desafio é encontrar colaboradores experientes ou mentores com quem interagir e aprender. Esse tipo de ajuda nem sempre é possível de se encontrar localmente. Isso exige uma escolha inicial entre trabalhar sozinho ou em colaboração com outros investigadores estabelecidos em outros locais.

Começando sozinho

Iniciar uma pesquisa sem a ajuda de um colaborador experiente é como ensinar a si mesmo a nadar: não é impossível, mas é muito difícil e às vezes permeado por perigos não antecipados. No entanto, em muitas situações, essa pode ser a única opção. Seguir algumas regras pode tornar o processo mais fácil.

- **Começar pelo mais simples.** Não é uma boa ideia começar a fazer pesquisa em uma comunidade por meio de um ensaio clínico randomizado. Pequenos estudos-piloto descritivos que produzem dados locais úteis podem ser mais vantajosos – é melhor alcançar um pequeno sucesso do que uma grande derrota. Projetos mais ambiciosos podem ser guardados para mais adiante e se basear nos dados do estudo-piloto realizado anteriormente. Por exemplo, um estudo descritivo sobre o uso de preservativos pelos jovens em Uganda conduzido por um pesquisador iniciante foi o primeiro passo para um amplo ensaio clínico de intervenção na prevenção de HIV/AIDS nessa comunidade (9).

- **Analisar as vantagens comparativas que o local oferece.** Quais questões podem ser melhor respondidas nesta comunidade do que em qualquer outro local? Isso, em geral, significa deixar o desenvolvimento de novas técnicas laboratoriais e tratamentos a cargo dos centros acadêmicos e de organizações de pesquisa farmacêutica. É melhor para um investigador iniciante se concentrar em problemas de saúde ou em populações incomuns em outros lugares, mas comuns na comunidade local.

- **Desenvolver redes.** Conforme foi discutido no Capítulo 2, uma rede de contatos é fundamental para qualquer pesquisador. Um jovem investigador deve estar sempre interagindo com outros cientistas que estejam abordando questões de pesquisa similares. Caso não seja possível obter apoio de colaboradores formais, talvez seja possível, pelo menos, que o investigador encontre alguém disposto a ler e a opinar sobre o esboço do protocolo de pesquisa, questionário ou manuscrito por *e-mail* ou telefone. Participar de congressos científicos na área de interesse é uma boa maneira de estabelecer contatos, e comentar o trabalho de um colaborador experiente pode ser uma boa maneira de iniciar um contato.

Pesquisa colaborativa

Já que começar sozinho é difícil, uma boa maneira de iniciar uma pesquisa em uma comunidade é colaborar com pesquisadores mais experientes sediados em outro local, especialmente se esses pesquisadores já estabeleceram confiança, contatos e metodologias no país de interesse. Existem dois modelos básicos para esse tipo de colaboração: de cima para baixo (*top-down*) e de baixo para cima (*bottom-up*) (10).

O modelo **de cima para baixo** refere-se aos estudos originados em um centro acadêmico e envolve investigadores comunitários no processo de recrutamento de pacientes e desenvolvimento do estudo. Isso ocorre, por exemplo, em ensaios clínicos multicêntricos de grande porte em que hospitais e serviços são convidados a arrolar pacientes para um protocolo de pesquisa já estabelecido. Essa abordagem tem as vantagens que advêm da colaboração com pesquisadores experientes, que são geralmente

responsáveis pelo delineamento do estudo e pela obtenção dos recursos e liberações necessários para conduzi-lo.

No modelo **de baixo para cima**, investigadores já estabelecidos fornecem orientação e assistência técnica a investigadores locais e a comunidades na elaboração de suas próprias agendas de pesquisa. Alguns centros médicos acadêmicos oferecem programas de treinamento para investigadores comunitários ou pesquisadores internacionais. Se um investigador conseguir obter acesso a um desses programas, ou estabelecer uma relação equivalente, isso pode ser uma oportunidade ideal para construir capacidade de pesquisa local, especialmente quando tal parceria é mantida a longo prazo. Entretanto, estabelecer relações institucionais desse tipo não é fácil. A maioria das agências está mais interessada em financiar projetos de pesquisa específicos do que em direcionar recursos para promover a capacidade de pesquisa local e colaborações. Mesmo quando há verba disponível para cobrir as despesas com treinamento e viagens, os investigadores experientes podem preferir usar seu tempo em suas próprias pesquisas a ajudar os outros a começar. α¡nda assim, nunca é demais destacar o valor da **pesquisa colaborativa e participativa de base comunitária** (*community-based participatory research*, CBPR), na qual a comunidade participa integralmente em todos os aspectos da pesquisa e que traz grandes benefícios em termos de satisfação, importância e relevância à comunidade local (11).

Os investigadores comunitários devem se valer das vantagens que podem oferecer aos investigadores mais experientes com quem desejam trabalhar. No modelo de cima para baixo, o melhor que podem oferecer é o acesso aos sujeitos de pesquisa. No modelo de baixo para cima, as vantagens podem incluir também o mérito científico intrínseco a um estudo comunitário, a coautoria em publicações decorrentes e a satisfação em construir uma relação de colaboração e auxiliar uma comunidade a desenvolver capacidade de pesquisa.

Para dar início a um novo programa de pesquisa, o ideal é formar uma parceria a longo prazo com uma instituição de pesquisa conceituada. **Memorandos de entendimentos** podem ser assinados pelas agências colaboradoras, permitindo fornecer aos potenciais financiadores evidências escritas de que houve comunicação e acordo. Estabelecer essa

colaboração de antemão pode poupar tempo e frustrações. Nessas colaborações pode haver uma combinação de projetos de cima para baixo e de baixo para cima. É importante lembrar que boas colaborações em pesquisa ocorrem fundamentalmente entre investigadores individuais. Uma instituição acadêmica pode fornecer o ambiente, a estrutura e os recursos para apoiar tais colaborações individuais, mas cabe aos pesquisadores envolvidos garantir a sensibilidade cultural, o respeito mútuo, o trabalho árduo e o compromisso a longo prazo para fazê-las funcionar.

PESQUISA INTERNACIONAL

Pesquisas internacionais envolvem a colaboração de diferentes grupos com níveis de experiência e recursos variados e, consequentemente, abrangem temas semelhantes aos de pesquisas comunitárias. Entretanto, esse tipo de pesquisa está sujeito a **desafios adicionais**. Os assuntos abordados na seção a seguir são de especial importância.

Barreiras de distância, idioma e cultura

Sem uma compreensão abrangente das perspectivas culturais de uma comunidade, até mesmo os planos mais bem elaborados podem fracassar, mesmo tendo sido realizado planejamento cuidadoso e utilizadas tecnologias avançadas. Para evitar que isso ocorra, os pesquisadores devem compreender as percepções culturais da doença nas comunidades onde pretendem trabalhar e desenvolver abordagens culturalmente sólidas para sua pesquisa colaborativa. Devido às **distâncias** envolvidas, as oportunidades para **comunicação** face a face entre colegas de diferentes países podem ser bastante limitadas. Quando possível, os colaboradores de ambos os países devem fazer pelo menos uma **visita** à outra instituição. Congressos internacionais às vezes oferecem oportunidades adicionais para um encontro, mas essas oportunidades não são muito frequentes. Felizmente, o *e-mail*, a *internet* e o *Skype* tornaram a comunicação internacional mais fácil, rápida e barata. Estabelecer uma boa comunicação atualmente é possível, mesmo a longa distância, entretanto, requer tempo e esforços especiais de ambos os lados. De nada adianta ter meios de comunicação modernos se eles não forem usados regularmente.

A falta de comunicação frequente e de respostas imediatas de cada lado é um sinal de que a colaboração à distância pode estar apresentando problemas.

As diferenças de **idiomas** geralmente se sobrepõem às barreiras causadas pela distância. Quando todos os investigadores envolvidos no estudo não falam a mesma língua materna, é importante estabelecer um idioma comum que todos possam falar, que geralmente é o inglês. Contudo, tornar o inglês a língua comum para todas as interações pode colocar os investigadores de muitos países em desvantagem.

É improvável que investigadores estrangeiros que não falam o idioma local tenham mais do que um entendimento superficial da cultura de um país e, portanto, não podem participar efetivamente em muitos aspectos importantes do estudo, incluindo a elaboração e a validação de questionários. Eles também terão dificuldade para conversar com os sujeitos e assistentes da pesquisa. Conseguir realizar essa comunicação é de suma importância em estudos que incluem componentes comportamentais.

Mesmo quando as barreiras linguísticas são superadas, as diferenças **culturais** α¡nda podem gerar mal-entendidos graves entre investigadores e sujeitos ou até mesmo entre diferentes investigadores. Traduções literais de questionários podem ter diferentes significados, ser culturalmente inapropriadas ou omitir fatores locais importantes. Normas institucionais podem ser diferentes. Por exemplo, em determinados contextos, o chefe do departamento do colaborador estrangeiro que teve pouco envolvimento direto no estudo poderia esperar que seu nome figurasse como primeiro autor na publicação decorrente. Tais questões devem ser discutidas previamente, e as decisões devem ser documentadas como parte do processo de desenvolvimento institucional para o projeto. Paciência, boa vontade e flexibilidade de todos os lados geralmente superam problemas desse tipo. Para projetos de maior porte, pode ser aconselhável incluir na equipe um antropólogo, um especialista em ética ou algum outro profissional com experiência em questões culturais.

A comunicação clara, aberta e frequente, assim como o pronto esclarecimento de quaisquer dúvidas ou questões são essenciais. Ao lidar com diferenças culturais e linguísticas, é melhor ser repetitivo e arriscar o

óbvio do que fazer suposições incorretas sobre o que os outros pensam ou dizem. Acordos de parceria por escrito que explicitam as responsabilidades e obrigações de ambos os lados podem ajudar a tornar mais claras questões como posse dos dados, ordem de autoria, direitos de publicação e decisões relacionadas a como apresentar os resultados da pesquisa. O desenvolvimento desses acordos requer atenção cuidadosa de ambas as partes envolvidas.

Questões sobre financiamento

Devido às desigualdades econômicas, as colaborações entre instituições de países ricos e países pobres geralmente são possíveis apenas com **financiamento** oriundo de doadores do país rico ou, com menor frequência, com fundos de outros países também ricos ou de organizações internacionais. Cada vez mais instituições de financiamento de grande porte estão se tornando ativas na pesquisa global em saúde, mas apoiam muitas vezes apenas agendas de pesquisa muito específicas, com exigências rigorosas para resultados mensuráveis. Grande parte dos recursos de financiamento bilateral flui pela instituição no país rico, reforçando a posição de subordinação da instituição dos países de baixa e média renda. Como em qualquer situação onde há **assimetria de poder**, isso gera desafios éticos. Quando os investigadores dos países ricos controlam os gastos, não é incomum que eles tratem seus companheiros dos países pobres mais como empregados do que realmente como colegas. Doadores internacionais e agências de fomento devem ter cuidado especial a fim de desencorajar esse tipo de postura, promovendo uma verdadeira parceria de atividades colaborativas (8).

As práticas diferentes de **gerenciamento financeiro** são também uma área de potencial conflito entre membros de consórcios de pesquisa. As instituições de países ricos podem tentar impor padrões de contabilidade difíceis ou impossíveis de serem alcançados localmente. As instituições de países de baixa e média renda podem cobrar encargos extras ou adicionar itens ao orçamento de pesquisa, tais como computadores e outros equipamentos com que pretendem ficar após o estudo. Esse fato é compreensível considerando-se as necessidades desses países e a inexistência de verbas alternativas. Entretanto, é importante que quaisquer subsídios além dos custos de pesquisa sejam negociados claramente e que

as práticas de contabilidade sejam implementadas de modo a respeitar as exigências das agências de financiamento. Por outro lado, como nos países de onde vem o financiamento os repasses institucionais e os salários costumam ser maiores, isso gera uma situação de desigualdade na medida em que a maior parte do financiamento para a pesquisa colaborativa fica nesses locais, mesmo quando o grosso do trabalho é realizado no país da instituição parceira.

As instituições de pesquisa e de financiamento do país onde se originou o financiamento devem prestar atenção especial à construção da **capacidade administrativa** em pesquisa dos parceiros locais. Isso pode significar oferecer treinamento administrativo e financeiro ou usar consultores especialistas nesses assuntos para auxiliar nas tarefas administrativas locais. Nos Estados Unidos, uma exigência para os parceiros internacionais é obter um Número D-U-N-S (Sistema Universal de Numeração de Dados, do inglês *Data Universal Numbering System*), que é um identificador único de nove dígitos para cada localização física de instituições que concorrem a editais de financiamento do governo norte-americano (http://fedgov.dnb.com/webform). Os esforços despendidos no desenvolvimento de capacidade administrativa podem ser recompensados com uma maior agilidade para cumprir prazos, maior eficiência no relato dos resultados, prevenção de conflitos desnecessários e construção de uma infraestrutura sólida para pesquisas futuras.

Questões éticas

A pesquisa internacional gera **questões éticas** que devem ser encaradas abertamente. Todas as questões éticas levantadas no Capítulo 14 valem nesse contexto. Por apresentar riscos especiais de violações da proteção de sujeitos humanos, esse tipo de pesquisa requer considerações e precauções adicionais.

Qual é, por exemplo, o **grupo de comparação apropriado** para se testar um novo tratamento em um país de baixa ou média renda onde o tratamento convencional não está disponível? É consensual que o uso de controle-placebo é antiético quando tratamentos eficazes já se tornaram padrão em outros lugares. Contudo, o que significa “atendimento-padrão” em uma comunidade onde a maioria da população não tem condições financeiras de usar tratamentos já comprovados e que podem estar

disponíveis em muitos países? Por um lado, os investigadores nem sempre podem oferecer um tratamento de ponta a todos os participantes de um estudo. Por outro lado, autorizar o uso de controle-placebo simplesmente devido ao acesso inadequado a medicamentos e cuidados médicos é antiético e vem sendo combatido por muitos grupos intergovernamentais e organizações que advogam pelos direitos das pessoas. Essas questões podem ser ilustradas por estudos sobre tratamentos antirretrovirais orais mais baratos para prevenir a transmissão vertical do HIV realizados em países onde a maioria das mulheres não tinham acesso a um tratamento já comprovado e disponível (12, 13).

Um problema relacionado a esse tema é a **testagem de tratamentos que provavelmente não serão economicamente acessíveis à população do país anfitrião**. Será que tais estudos são éticos mesmo que sigam as regras usuais? Por exemplo, seria ético estudar um novo medicamento para o diabetes tipo 2 em um país de baixa ou média renda onde esse medicamento provavelmente teria custo proibitivo? As respostas a essas questões não são simples. As convenções internacionais que governam a pesquisa ética, como a Declaração de Helsinque, têm sido questionadas e são sujeitas a muitas interpretações (14,15).

A pergunta-chave talvez seja, primeiramente, considerar por que o estudo está sendo conduzido em um país de baixa ou média renda. Se o verdadeiro objetivo for coletar informações que possam ajudar a população do país, isso deve contar a favor do estudo, que deve ser planejado de modo a atingir esse fim. Idealmente, o objetivo da pesquisa deveria ser uma mudança sustentável e um valor agregado para o país anfitrião (16). Se, por outro lado, o objetivo for a conveniência ou simplesmente evitar os obstáculos que seriam encontrados se o estudo fosse realizado em um país rico, então esse estudo deveria ser submetido a todos os requerimentos éticos que seriam, de outro modo, necessários no país de quem promove o estudo, incluindo a importante exigência da justiça distributiva (ver Capítulo 14).

Por essa e outras razões, os estudos conduzidos em países pobres cujos recursos financeiros são providos de fora devem ser aprovados pelos **CEPs de ambos os países**. Entretanto, embora essa aprovação seja necessária, ela não garante que o estudo seja ético. Os órgãos de revisão ética em muitos países pobres são pouco desenvolvidos ou até mesmo

inexistentes e, em alguns casos, podem ser manipulados por investigadores locais ou políticos. Por outro lado, os CEPs dos países ricos às vezes são insensíveis ou até mesmo ignoram questões especiais inerentes à pesquisa internacional. A aprovação oficial não isenta os investigadores da responsabilidade final pela condução ética da pesquisa.

Outra preocupação ética importante é o **tratamento dos colaboradores** de países de baixa ou média renda. Há várias questões que devem ser estabelecidas previamente. Quem será o responsável pelos dados que serão gerados? Quem precisa de permissão e quem dará essa permissão para conduzir e publicar análises? Os investigadores dos países pobres terão o apoio necessário para elaboração de manuscritos para publicações internacionais sem ter de abrir mão da primeira autoria? Qual a duração do compromisso sendo firmado por ambos os lados? Por exemplo, um ensaio clínico recente de grande porte em vários países pobres sobre aconselhamento e exame voluntários na prevenção de infeção por HIV foi abruptamente encerrado na Indonésia (17). De acordo com os investigadores, isso se deveu ao fato de que o desfecho de interesse foi menos comum naquele local do que o previsto nos cálculos de poder estatístico do estudo. Apesar de ter sido uma decisão de senso prático, para os indonésios isso significou uma quebra de confiança.

Outras questões éticas podem ter a ver com as **realidades políticas e econômicas locais**. Por exemplo, um ensaio clínico planejado para investigar a profilaxia pré-exposição ao HIV com tenofovir para prostitutas foi cancelado, embora tenha sido aprovado pelos comitês de ética de todos os países envolvidos (18). Os sujeitos que iriam participar do estudo ficaram preocupados com a possibilidade de não terem atendimento médico para os problemas relacionados à infecção pelo HIV ou aos efeitos adversos da medicação e se recusaram a participar sem a garantia de seguro de saúde vitalício. O primeiro ministro do país interveio para encerrar o estudo.

Por fim, toda colaboração internacional deveria ter como objetivo explícito a melhoria da **capacidade local de pesquisa**. Que habilidades e equipamentos o projeto deixará no local após seu término? Que atividades de treinamento serão viabilizadas àqueles que trabalham no projeto? Os participantes locais terão a oportunidade de participar de congressos internacionais? Se sim, isso valerá apenas para os pesquisadores seniores

que já usufruem de tais oportunidades ou seus colegas juniores também terão a chance de participar? Os pesquisadores locais serão realmente colaboradores e autores principais dos artigos ou simplesmente empregados na coleta de dados? Os cientistas dos países pobres devem questionar e exigir respostas claras para todas essas questões. Como apresentado na Tabela 18.2, uma **comunicação efetiva e um compromisso a longo prazo** são temas recorrentes nas pesquisas colaborativas internacionais bem-sucedidas.

TABELA 18.2 Estratégias para fortalecer a pesquisa colaborativa internacional

Cientistas de países de baixa e média renda
Escolher os colaboradores com cuidado
Aprender inglês (ou outro idioma dos colaboradores)
Familiarizar-se com a literatura científica internacional da área de estudo
Certificar-se de que a colaboração proporcionará a construção de capacidade local de pesquisa
Esclarecer as expectativas científicas e administrativas com antecedência
Cientistas de países ricos
Escolher os colaboradores com cuidado
Aprender a cultura e o idioma local
Ser sensível às questões éticas locais
Incentivar a participação de colaboradores locais em todos os aspectos do processo de pesquisa
Esclarecer as expectativas científicas e administrativas com antecedência
Agências de fomento à pesquisa
Estabelecer prioridades de recursos com base nas necessidades de saúde pública
Incentivar a colaboração genuína, em vez do modelo “de cima para baixo”
Reconhecer a importância da construção de capacidade local de pesquisa
Tornar os subsídios para equipamentos e infraestrutura locais explícitos
Assegurar que os repasses institucionais e os altos salários no país rico não tomem muito do orçamento

A Organização Mundial da Saúde recentemente publicou um conjunto de estudos de caso que lidavam com questões éticas na pesquisa sobre saúde global (19), para auxiliar investigadores, membros de comitês de ética, autoridades em saúde e outros a desempenharem seus papeis respectivos na condução ética da pesquisa.

Muito se pode aprender a partir dos erros e sucessos dos outros, mas com boa vontade por parte dos financiadores, dos parceiros dos países de onde se origina o financiamento e dos participantes de ambos os lados da parceria de pesquisa, é possível contemplar os princípios éticos na pesquisa internacional, fortalecendo globalmente a capacidade de se realizar esse tipo de pesquisa.

Riscos e frustrações

Pesquisadores de países ricos que almejam tomar parte em pesquisa

internacional precisam iniciar com uma avaliação realista das dificuldades e dos riscos envolvidos. Começar esse tipo de trabalho é, geralmente, um processo longo e vagaroso, com muitos **obstáculos burocráticos** em ambos os países. Nos países onde não há infraestrutura e estabilidade política, anos de trabalho podem ser vulneráveis a **catástrofes** naturais ou humanas. Em casos extremos, essas situações podem ameaçar a segurança da equipe de pesquisa ou dos investigadores. Por exemplo, programas importantes de pesquisa colaborativa em HIV/AIDS construídos ao longo de muitos anos foram completamente destruídos devido às guerras civis na Ruanda e no Congo.

Desafios menos catastróficos, porém mais comuns, são as dificuldades diárias e os **riscos à saúde** enfrentados por investigadores que trabalham fora de seus países, podendo variar desde água contaminada e malária até *smog*, crimes comuns e acidentes de trânsito.

Outra frustração para pesquisadores em países de baixa ou média renda é a dificuldade em **aplicar seus achados**. Mesmo quando novas estratégias de prevenção ou tratamento de uma doença podem ser desenvolvidas com sucesso e são comprovadamente eficazes, a falta de vontade política ou de recursos pode ser um empecilho para sua ampla aplicação no país onde foi realizada a pesquisa. Os pesquisadores precisam ser realistas quanto às suas expectativas e direcionar seu trabalho para a investigação de estratégias que sejam passíveis de implementação no caso de serem consideradas eficazes. Além disso, devem estar preparados para advogar pela melhoria da saúde da população em estudo.

As recompensas

Apesar das dificuldades, há uma necessidade pungente de mais pesquisas em saúde em muitas partes do mundo. Ao participar de uma pesquisa internacional, um investigador de um país doador pode exercer um **impacto** muitas vezes maior e mais imediato em termos de **saúde pública** do que faria se mantivesse suas atividades entre quatro paredes na sua instituição acadêmica. Esse impacto emana não apenas da própria pesquisa, mas também daquilo que é muitas vezes denominado **diplomacia em saúde global**. De fato, a saúde é agora vista como uma grande força motriz para as prioridades de políticas de relações exteriores

(20). A diplomacia em saúde pode ser praticada por meio de pesquisa colaborativa sobre os desafios globais de saúde, como HIV/AIDS, malária, tuberculose, saúde materno-infantil e fortalecimento de sistemas de saúde. A saúde e a política sempre estiveram interconectadas, porém, em um mundo globalizado, há uma necessidade crescente de ações colaborativas para abordar os principais problemas de saúde que ultrapassam as fronteiras dos países; a pesquisa internacional é parte desse esforço global. A oportunidade de ter uma participação genuína e fazer contribuições realmente significativas para a saúde global é um privilégio que pode **enriquecer a carreira** e a vida pessoal daqueles envolvidos. Todos podem ganhar por meio de mais colaboração e pela expansão das oportunidades de pesquisa.

RESUMO

1. A **pesquisa comunitária e internacional** é necessária para detectar **diferenças regionais** em questões como **epidemiologia** de uma doença, **fatores culturais** e outros **fatores locais** que determinam quais **intervenções** serão **efetivas**.
2. A **participação local** na pesquisa clínica pode trazer benefícios secundários para a região, como maior **capacidade acadêmica** e **autossuficiência**.
3. Embora os aspectos teóricos e éticos da pesquisa comunitária e internacional sejam amplamente aplicáveis, os aspectos práticos, como a obtenção de recursos financeiros ou a identificação de um mentor, são **mais difíceis** nesse contexto; estratégias de sucesso incluem começar com **estudos pequenos**, pensar sobre as **vantagens locais** e **construir redes**.
4. A **colaboração** entre centros médicos acadêmicos e **pesquisadores comunitários** pode seguir um modelo de **cima para baixo** (os investigadores comunitários conduzem estudos originados no centro acadêmico) ou um modelo de **baixo para cima** (os investigadores dos centros acadêmicos auxiliam os investigadores comunitários a conduzir estudos que eles mesmos originam).
5. A **pesquisa internacional** tem características semelhantes às da pesquisa comunitária, porém apresenta **desafios** adicionais,

especialmente em **países de baixa e média renda**, relacionados a **comunicação** e **idioma**, **diferenças culturais**, **recursos financeiros**, **assimetria nas estruturas de poder** e **práticas administrativas** e **financeiras**.

6. A pesquisa internacional tem **questões éticas** próprias, incluindo **tratamentos com custo inacessível** em países de baixa ou média renda, uso de **placebos** em **populações vulneráveis** e a questão do *status* e tratamento dos **colaboradores**.
7. Superar os desafios da pesquisa internacional traz gratificações, como a de **ajudar pessoas necessitadas**, fazer parte de uma **comunidade de saúde global mais ampla** e **vivenciar experiências interculturais** enriquecedoras.

REFERÊNCIAS

1. Unite for Sight. The importance of global health research: closing the 10/90 gap. Available at: http://www.uniteforsight.org/global-impact-lab/global-health-research#_ftnref12, accessed 9/23/12.
2. Hearst N, Chen S. Condom promotion for α_{i}DS prevention in the developing world: is it working? *Studies in Family Planning* 2004;35(1):39–47.
3. Drugs for hypertension. *Med Lett Drugs Ther* 1999;41:23–28.
4. Griffith BN, Lovett GD, Pyle DN, et al. Self-rated health in rural Appalachia: health perceptions are incongruent with health status and health behaviors. *BMC Public Health* 2011;11:229. doi:10.1186/1471-2458-11-229.
5. Nutting PA, Beasley JW, Werner JJ. Practice-based research networks answer primary care questions. *JAMA* 1999;281:686–688.
6. Miller RS, Ivenson DC, Fried RA, et al. Carpal tunnel syndrome in primary care: a report from ASPN. *J Fam Pract* 1994;38:337–344.
7. United Nations Department of Economic and Social Affairs (UN DESA). Trends in international migrant stock: the 2008 revision. Available at: *http://esa.un.org/migration/index.asp?panel=1*, accessed 1/12/2013.
8. Lee K, Mills A. Strengthening governance for global health research: the countries that most need health research should decide what should be funded. *BMJ* 2009;2000:775–776.
9. Kajubi P, Kamya MR, Kamya S, et al. Increasing condom use without reducing HIV risk: results of a controlled community trial in Uganda. *Journal of* α_{i}*DS* 2005;40(1):77–82.
10. Hearst N, Mandel J. A research agenda for α_{i}DS prevention in the developing world. *AIDS* 1997;11(Suppl 1):S1–4.
11. Minkler M and Wallerstein N, eds. (2008). *Community-Based Participatory*

Research for Health: From Process to Outcomes. ISBN 978-0-470-26043-2. Jossey-Bass

12. Lurie P, Wolfe SM. Unethical trials of interventions to reduce perinatal transmission of the human immunodeficiency virus in developing countries. *N Engl J Med* 1997;337:853–856.
13. Perinatal HIV Intervention Research in Developing Countries Workshop Participants. Science, ethics, and the future of research into maternal-infant transmission of HIV-1. *Lancet* 1999;353:832–835.
14. Brennan TA. Proposed revisions to the Declaration of Helsinki: will they weaken the ethical principles underlying human research? *N Engl J Med* 1999;341:527–531.
15. Levine RJ. The need to revise the Declaration of Helsinki. *N Engl J Med* 1999;341:531–534.
16. Taylor D, Taylor CE. *Just and lasting change: when communities own their futures.* Baltimore: JHU Press, 2002.
17. Kamenga MC, Sweat MD, De Zoysa I, et al. The voluntary HIV-1 counseling and testing efficacy study: design and methods. *AIDS and Behavior* 2000;4:5–14.
18. Page-Shafer K, Saphonn V, Sun LP, et al. HIV prevention research in a resource-limited setting: the experience of planning a trial in Cambodia. *Lancet* 2005;366(9495):1499–1503.
19. Cash R, Wikler D, Saxena A, et al. *Casebook on ethical issues in international health research*. Geneva: World Health Organization, 2009.
20. Katz R, Kornblet S, Arnold G, et al. Defining health diplomacy: changing demands in the era of globalization. *The Milbnk Quarterly* 2011;89(3):503–523.

CAPÍTULO 19

Redigindo uma proposta para solicitar financiamento de pesquisa

Steven R. Cummings, Deborah G. Grady e Stephen B. Hulley

O protocolo é o plano detalhado do estudo. Escrever o protocolo força o investigador a organizar, a esclarecer e a aperfeiçoar todos os elementos da pesquisa, aumentando seu rigor científico e sua eficiência. Mesmo quando não há necessidade de solicitar financiamento, ainda assim é fundamental elaborar um protocolo para conduzir o trabalho e para obter a aprovação do comitê de ética em pesquisa (CEP). A **proposta** (*proposal*) é um documento que é redigido para solicitar financiamento de pesquisa. Ela contém uma descrição dos objetivos do estudo, da sua relevância, da sua metodologia, das preocupações relacionadas aos participantes, do orçamento e de outras informações administrativas e de apoio exigidas especificamente pela agência de financiamento.

Este capítulo descreve **como redigir uma proposta** com maior chance de **obter financiamento**. Ele concentra-se em propostas de pesquisas originais utilizando o formato sugerido pelo National Institutes of Health (NIH) dos Estados Unidos, porém as propostas para a maioria das outras agências de financiamento geralmente seguem um formato parecido. No site do NIH (http://grants.nih.gov/grants/writing_application.htm) estão disponíveis excelentes orientações sobre como redigir uma proposta, preparar o orçamento e realizar a submissão.

REDIGINDO UMA PROPOSTA

A tarefa de redigir uma proposta geralmente exige vários meses de organização, de redação e de revisão. Os passos a seguir podem ajudar a deslanchar o projeto com sucesso.

- **Decidir onde a proposta será submetida.** Cada agência de financiamento tem suas próprias áreas de interesse, processos e procedimentos para elaboração de propostas de pesquisa. Por isso, o investigador deve decidir previamente onde submeter sua proposta, verificar os limites especificados para financiamento, obter **instruções específicas** sobre como elaborar a proposta e informar-se sobre os prazos exigidos pela agência escolhida. O site do NIH na internet é um bom ponto de partida (http://grants.nih.gov/grants/oer.htm). As áreas de interesse podem ser identificadas por meio dos sites de cada instituto, nas seções em que são descritas as prioridades daquele instituto. Informações adicionais sobre as áreas atuais de interesse podem ser obtidas **conversando** com administradores científicos dos institutos do NIH, cujas informações para contato e áreas de responsabilidade são listadas na seção de Anúncios de Oportunidades para Financiamento do NIH (NIH *Funding Opportunity Announcements*) e nas páginas na internet dos institutos.

- **Organizar uma equipe e designar um líder.** A maioria das propostas é redigida por uma equipe constituída de várias pessoas que irão conduzir o estudo. Essa equipe pode ser pequena (somente o investigador e seu mentor) ou grande (incluindo colaboradores, um bioestatístico, um administrador financeiro, assistentes de pesquisa e o pessoal de apoio). É importante que a equipe detenha o conhecimento especializado necessário (ou tenha como obtê-lo) para delinear o estudo.

  Um membro da equipe deve assumir a responsabilidade de liderar o trabalho. Esse indivíduo será o **investigador principal** (IP), que terá, em última instância, a autoridade máxima e a responsabilidade pelo estudo. O IP deve exercer liderança durante todo o processo de desenvolvimento da proposta, delegando responsabilidades pela redação e outras tarefas, estabelecendo prazos, conduzindo reuniões periódicas da equipe, assegurando que todas as tarefas estejam concluídas em tempo hábil e cuidando pessoalmente da qualidade da

proposta.

Geralmente o IP é um cientista experiente cujo conhecimento e sabedoria possam ser úteis nas decisões de delineamento e cujo montante de estudos prévios aumente a probabilidade de sucesso do estudo e, dessa forma, de se obter o financiamento solicitado. Dito isso, deve-se destacar que o NIH encoraja **investigadores iniciantes** a submeterem propostas como IPs, dispondo de editais especiais para eles, e frequentemente dando prioridade para financiar suas propostas (http://grants.nih.gov/grants/new_investigators/). A definição do NIH para investigador iniciante é um cientista que ainda não foi IP de um auxílio de pesquisa do NIH. Porém, os **IPs estreantes** terão maior probabilidade de conseguirem o financiamento se já tiverem experiência prévia em pesquisa – sob a supervisão de um pesquisador sênior e com financiamento fornecido por aquele indivíduo, por um auxílio para desenvolvimento de carreira ou por pequenos auxílios institucionais ou de fundações. Já ter **publicado** artigos, inclusive como primeiro autor, é essencial para fornecer evidências de que o investigador iniciante tem potencial para ser um cientista independente bem sucedido e está preparado para liderar a pesquisa.

Um IP estreante deve incluir na proposta de auxílio **co-investigadores** que tenham experiência consagrada na área de interesse, para fornecer orientação sobre a condução do estudo e para aumentar a chance de uma avaliação favorável. Às vezes, isso pode ser feito utilizando o mecanismo de **múltiplos IPs**; o NIH permite mais de um investigador caso cada um deles traga expertises diferentes, porém complementares, e seus papéis e responsabilidades distintos sejam claramente definidos (http://grants.nih.gov/grants/multi_pi/overview.htm).

- **Seguir as diretrizes da agência de financiamento.** As instituições de fomento à pesquisa fornecem por escrito **diretrizes** que devem ser analisadas cuidadosamente pelos investigadores antes de se dar início à redação da proposta. Elas incluem os tipos de pesquisa que serão financiados e instruções detalhadas para a elaboração da proposta, o número máximo de páginas, os limites permitidos para financiamento, o cronograma e outros elementos que deverão fazer parte da proposta.

Entretanto, essas diretrizes não fornecem todas as informações necessárias sobre a maneira como as agências operam e suas preferências. No estágio inicial de elaboração das propostas, é interessante que o investigador **discuta seu plano** com alguém ligado à agência, que poderá esclarecer as preferências daquela instituição (como escopo e grau de detalhamento exigidos para a proposta) e confirmar se o plano de pesquisa satisfaz os interesses desta. O NIH, outras agências federais e as fundações privadas em geral contam com **administradores científicos** cuja função é assessorar os investigadores na adequação de suas propostas às prioridades de financiamento das agências. Pode ser muito útil contatar por e-mail ou telefone o administrador científico responsável pela área de pesquisa de interesse, para esclarecer as diretrizes da agência, seus interesses e procedimentos de revisão. Posteriormente, encontrá-lo em algum congresso onde ambos estarão presentes ou visitá-lo quando o investigador viajar para algum lugar próximo da sede da agência é uma boa forma de estabelecer uma relação produtiva que facilita a elaboração de propostas com maiores chances de financiamento.

É muito útil elaborar uma *checklist* contendo os detalhes exigidos e checá-los repetidas vezes antes de submeter a proposta. A rejeição de uma proposta de qualidade por falta de adequação aos detalhes especificados é uma experiência frustrante e que pode ser evitada. Os gerentes de financiamento das universidades geralmente dispõem de *checklists* que eles revisam antes de submeter uma proposta.

- **Estabelecer um cronograma e fazer reuniões periódicas.** Um cronograma com as datas-limite para finalização das tarefas pressiona os membros da equipe a cumprirem suas obrigações no prazo planejado. Além de conter os componentes científicos especificados pela agência, o **cronograma** deverá prever as exigências administrativas da instituição onde a pesquisa irá ocorrer. As universidades em geral requerem revisão demorada do orçamento e dos subcontratos antes de submeter a proposta à agencia financiadora, de forma que o prazo real para concluir a proposta pode ser vários dias ou mesmo semanas antes do prazo da agência. Negligenciar esses detalhes pode gerar atropelos de última hora, pondo em risco uma proposta até

então bem elaborada.

Um cronograma funciona melhor quando explicita datas-limite para produções escritas e quando os membros do grupo participam da elaboração de suas próprias tarefas. O cronograma deve ser revisado em encontros periódicos presenciais ou virtuais para verificar se as tarefas estão em dia e se as datas-limite ainda são viáveis.

- **Procurar propostas-modelo.** Pode ser de grande proveito buscar **propostas aprovadas recentemente** pela mesma agência onde o financiamento está sendo requerido. Essa proposta ilustra de maneira concreta o conteúdo e o formato de uma boa proposta. O investigador poderá adaptar as melhores ideias do modelo, elaborando uma proposta mais clara, lógica e persuasiva. É útil também ter em mãos alguns pareceres críticos fornecidos pela agência para propostas que foram submetidas, tenham sido elas aprovadas ou não, pois ilustram os pontos considerados relevantes aos cientistas que farão a revisão da proposta. Esses exemplos podem ser obtidos de colegas ou do setor de pesquisas financiadas da instituição do investigador.

- **Trabalhar a partir de um roteiro.** Começar redigindo a proposta a partir de um roteiro (Tabela 19.1) ajuda a organizar as tarefas a serem cumpridas. Se várias pessoas estiverem trabalhando juntas, o roteiro auxilia a **distribuir as responsabilidades** na preparação da proposta. Um dos obstáculos mais comuns ao se elaborar um roteiro é a ideia de que o plano de pesquisa deva estar montado, antes mesmo de se escrever a primeira frase. É importante deixar as ideias fluírem livremente para o papel, criando-se uma matéria-prima para posterior edição, aperfeiçoamento e aconselhamento específico de colegas.

- **Rever e revisar repetidas vezes.** Escrever uma proposta é um **processo iterativo** que resulta em muitas versões, cada uma delas refletindo novas ideias, sugestões e dados adicionais. As primeiras versões da proposta devem ser revisadas criticamente por colegas familiarizados com o assunto e com a agência de fomento. Atenção especial deve ser dada à relevância e ao caráter inovador da pesquisa, à validade do delineamento e dos métodos e à clareza da redação. É melhor receber

críticas duras e detalhadas antes de a proposta ser submetida a vê-la rejeitada por não se terem antecipado e abordado problemas. Quando a proposta estiver praticamente pronta para ser submetida, o passo final será uma revisão cuidadosa, levando-se em conta consistência interna, formatação, cumprimento das exigências da agência, bem como correção ortográfica e gramatical. Uma redação desleixada passa a ideia de um trabalho desleixado e de uma liderança incompetente, desviando a atenção de ideias que são boas nos demais aspectos.

ELEMENTOS DE UMA PROPOSTA PARA UM GRANDE AUXÍLIO DE PESQUISA

Os elementos de uma proposta para um auxílio de pesquisa de grande porte como o NIH R01 são apresentados na Tabela 19.1. As propostas para outros tipos de auxílios e de contratos do NIH e de outras agências de financiamento podem requerer menos informações ou um formato diferente. O investigador deve prestar muita atenção às exigências da agência que irá recebê-la.

TABELA 19.1 Elementos principais de uma proposta, com base no modelo do NIH

- Título
- Resumo do projeto
- Partes administrativas
  - Orçamento e justificativa do orçamento
  - Currículos resumidos (biosketches) dos investigadores
  - Instalações e recursos
- Objetivos específicos
- Estratégia de pesquisa
  - Relevância
  - Caráter inovador
  - Abordagem
    - Quadro geral
    - Justificativa para a pesquisa planejada e dados preliminares
    - Sujeitos do estudo
      - Critérios de seleção
      - Delineamento da amostragem
      - Planos para o recrutamento
      - Planos para otimizar a adesão e o seguimento completo
    - Procedimentos do estudo (se aplicável)
      - Randomização

- Cegamento
- Aferições
  - Variáveis preditoras principais (intervenção, se ensaio clínico)
  - Potenciais variáveis confundidoras
  - Variáveis de desfecho
- Estatística
  - Abordagem para as análises estatísticas
  - Hipóteses, tamanho de amostra e poder estatístico
- Conteúdo e cronograma das visitas do estudo
- Gerenciamento de dados e controle de qualidade
- Cronograma e organograma
- Limitações e abordagens alternativas

Sujeitos humanos

Referências

Apêndices e acordos de colaboração

Início

O **título** deve ser descritivo e conciso. Ele fornece a primeira impressão e uma lembrança duradoura do objetivo geral e do delineamento do estudo. Por exemplo, o título “Ensaio clínico randomizado sobre ultrassom de alta frequência guiado por RM vs. ultrassom simulado para tratar miomatose uterina sintomática” sumariza de forma sucinta a questão de pesquisa e o delineamento. Devem-se evitar frases desnecessárias e vazias como “Um estudo para determinar o…”.

O **resumo do projeto** deve sumarizar o protocolo de forma concisa. Ele deve começar com os objetivos da pesquisa e sua justificativa, apresentar o delineamento e os métodos e concluir com o impacto esperado dos possíveis achados do estudo. O resumo deve ser informativo para pessoas que trabalham na mesma área ou em áreas relacionadas e compreensível para um leitor leigo porém com bom conhecimento científico. A maioria das agências exige que o resumo se limite a um certo número de palavras; portanto, é importante usar termos eficientes e descritivos. O resumo em geral é escrito depois que os outros elementos do protocolo estão estabelecidos e deve ser submetido a várias revisões até alcançar a excelência necessária. Essa será a única página lida por muitos dos revisores e será uma lembrança conveniente dos principais elementos da proposta para os demais leitores. Por esse motivo, o resumo deve falar por si só, incorporando todos os elementos e descrevendo de forma convincente os pontos fortes e os potenciais impactos do estudo proposto.

Seções administrativas

Geralmente, todas as agências de financiamento exigem uma seção administrativa que apresente o orçamento, uma descrição das qualificações do pessoal, dos recursos da instituição do investigador e do acesso a equipamentos, espaço e *expertise*.

A seção de **orçamento** é, em geral, organizada de acordo com as diretrizes da instituição financiadora. Propostas para o NIH, por exemplo, devem seguir um modelo preestabelecido apresentando um orçamento detalhado para os primeiros 12 meses e um orçamento resumido para todo o projeto (geralmente de 2 a 5 anos). O orçamento dos primeiros 12 meses deve incluir as seguintes categorias de gastos: pessoal (incluindo os nomes e as funções de todos os envolvidos no projeto, o percentual de tempo que cada um despenderá no projeto e o valor-moeda referente aos salários e encargos sociais para cada indivíduo); consultoria; equipamentos; materiais de consumo; viagens; gastos com cuidados de pacientes; alterações e renovações; contratos de serviços e outros gastos (p. ex., telefone, correspondência, conferências, fotocópias, ilustrações, publicações, compra de livros e prestações de serviços).

Não se deve deixar o orçamento para o último minuto. Muitos elementos requerem tempo (p. ex., estimar gastos com espaço, equipamentos e pessoal). As universidades geralmente dispõem de administradores experientes cujo trabalho é auxiliar os investigadores a preparar os orçamentos e outras partes administrativas de uma proposta. É aconselhável contatar esse administrador logo no início e planejar encontros ou reuniões telefônicas regulares para avaliar o progresso e o cumprimento do cronograma da seção administrativa. O administrador pode começar a trabalhar tão logo tenha sido elaborado o roteiro, recomendando as quantias necessárias para cada item do orçamento e ajudando a garantir que o investigador não deixe de considerar gastos importantes. As instituições possuem normas e prazos que devem ser cumpridos; assim, um administrador experiente poderá ajudar a prever regras da sua instituição e possíveis dificuldades e atrasos. O administrador também poderá ajudar a redigir a primeira versão das seções de justificativa do orçamento e recursos e a juntar os currículos, apêndices e outros materiais de apoio para a proposta.

Todos os itens do orçamento devem ser explicados detalhadamente em uma **justificativa de orçamento**. Os salários em geral abrangem a maior fatia dos gastos em um projeto de pesquisa clínica; por isso, é importante documentar a necessidade de cada indivíduo que consta na folha de pagamento e suas responsabilidades específicas, para justificar o percentual de trabalho que a ele será designado. A descrição das tarefas dos investigadores e dos outros membros da equipe deve ser completa, porém sucinta, não deixando dúvida aos revisores de que seu trabalho é essencial para o sucesso do projeto.

Os revisores costumam se preocupar com o tempo que membros importantes da equipe pretendem dedicar ao projeto. Muitas propostas são mal avaliadas porque membros importantes preveem um comprometimento muito pequeno no estudo em questão e um comprometimento maior em outras atividades, dando a entender que eles poderão não despender a energia necessária ao estudo proposto. Porém, é mais comum eles reagirem negativamente a percentuais que estejam muito além das exigências da tarefa descrita.

Mesmo se o **orçamento** for bem-elaborado, ele poderá sofrer **alterações** se mudarem as necessidades do estudo ou por outros motivos, como, por exemplo, gastos inesperados ou contenção de despesas. Em geral, uma vez concedida a verba, o investigador tem autonomia para gastá-la de outra forma que não a especificada no orçamento, desde que as mudanças sejam mínimas e os gastos sejam relacionados aos objetivos do estudo. Quando o investigador deseja remanejar recursos entre categorias ou fazer mudanças substanciais (para mais ou para menos) no percentual de dedicação de membros importantes da equipe, ele precisa obter o consentimento da agência financiadora. As agências geralmente aprovam pedidos razoáveis de remanejo de orçamento, desde que o investigador não peça aumento no valor total do financiamento.

O NIH exige um currículo resumido (*biosketch*) de todos os investigadores e consultores que serão financiados pelo auxílio. Esses **currículos resumidos** são resumos de quatro páginas seguindo um formato-padrão que inclui um relato pessoal sobre como a experiência do investigador o qualifica para conduzir o estudo e lista sua formação e treinamento, cargos que ocupou e empregos, prêmios, um número limitado de publicações relevantes e auxílios de pesquisa e contratos

relevantes.

A seção sobre os **recursos** disponíveis ao projeto pode incluir computadores, equipamentos técnicos e espaço para laboratórios e escritórios, acesso a exames de imagem e equipamentos para aferições, bem como recursos para facilitar o recrutamento dos participantes, a coleta e gerenciamento de dados e a estocagem das amostras. A seção sobre os recursos frequentemente incorpora descrições de outras propostas prévias ou de materiais fornecidos pela instituição, centro ou laboratório do investigador.

Objetivos específicos

Os objetivos específicos são formulações da questão de pesquisa utilizando termos concretos para especificar o desfecho desejado. Essa seção de uma proposta do NIH deve ser **concisa**, pois ela é restrita a uma página. E, uma vez que essa é a página à qual muitos revisores dedicam mais atenção, ela deve ser escrita com muito cuidado e revisada repetidas vezes na medida em que a proposta é desenvolvida.

Um padrão comum é iniciar com dois a três parágrafos curtos que sumarizam as informações básicas: questão de pesquisa e sua importância, estudos que já foram realizados e como eles não resolveram o problema e a abordagem que está sendo proposta para responder à questão de pesquisa neste estudo. Após, é apresentada uma breve formulação dos objetivos específicos, expressos como objetivos descritivos tangíveis e, sempre que possível, como **hipóteses** testáveis.

Os objetivos são apresentados em uma sequência lógica adequada ao estudo que está sendo planejado. Pode-se iniciar com os objetivos transversais para o período da linha de base, seguidos pelos objetivos relacionados aos achados do seguimento. Outra opção é iniciar com objetivos que abordam mecanismos fisiopatológicos e concluir com objetivos que abordam desfechos clínicos ou de saúde pública. Um padrão que funciona particularmente bem para bolsas de desenvolvimento de carreira (*career development awards*), denominado “pesquisa com métodos mistos” (*mixed methods research*) inicia com objetivos qualitativos que podem utilizar grupos focais para desenvolver um instrumento ou intervenção-chave, seguidos por objetivos quantitativos com variáveis preditoras e de desfecho e testes de hipóteses. Outro padrão

é iniciar com o objetivo mais importante, com o intuito de destacá-lo; a sequência de objetivos geralmente serve como **roteiro** para organizar as seções posteriores da proposta e, portanto, isso tem a vantagem de colocar o objetivo principal em primeiro lugar em todas as demais seções da proposta, como as relacionadas ao tamanho da amostra e ao poder estatístico.

A seção de Objetivos Específicos costuma terminar com um parágrafo curto que resume de forma concisa o potencial **impacto** dos achados do estudo sobre o conhecimento de saúde e doença, sobre a prática clínica, sobre a saúde pública ou sobre pesquisas futuras. A meta é construir uma proposta convincente que levará os membros do comitê que não foram os revisores primários ou secundários (e que poderão ter lido apenas essa única página) a defenderem uma excelente nota para a proposta.

Estratégia de pesquisa

O formato atual do NIH estabelece para a maioria dos tipos de propostas um limite de 12 páginas para apresentar a **estratégia de pesquisa**, que é dividida em três seções:

- A seção de **relevância**, em geral de duas a três páginas, descreve como os achados do estudo irão contribuir para avançar o conhecimento científico, para abordar um problema importante ou barreira ao progresso naquele campo, para melhorar a prática clínica ou de saúde pública ou para influenciar políticas de saúde. Esta seção pode resumir brevemente a magnitude do problema e o que já foi alcançado, definir os problemas com base no conhecimento atual e mostrar como o estudo proposto irá contribuir para avançar nessa área.
- A seção de **inovações**, geralmente de uma a duas páginas, aponta como o estudo proposto difere de pesquisas prévias sobre o assunto. Ela pode enfatizar o potencial para documentar novos mecanismos de doença ou o fato de o estudo utilizar novos métodos de aferição, populações diferentes ou maiores, novos métodos terapêuticos ou preventivos ou novas abordagens para analisar os dados. As diretrizes do NIH focam em como a pesquisa irá mudar os paradigmas atuais de pesquisa ou prática clínica por meio da utilização de conceitos, métodos ou

intervenções inovadores. Dito isso, muitos estudos clínicos que receberam financiamento resultam apenas em pequenas melhorias e refinamentos em conceitos, métodos ou intervenções. O nosso conselho é descrever as características inovadoras da pesquisa de forma acurada, sem exagerar que o estudo irá mudar paradigmas ou utilizar métodos inteiramente novos.

- A seção sobre a **abordagem** (anteriormente denominada "métodos") em geral ocupa de sete a nove páginas. Ela detalha o delineamento e a condução do estudo e é avaliada cuidadosamente pelos revisores. As diretrizes do NIH sugerem que a seção sobre abordagem seja organizada de acordo com os objetivos específicos e que inclua componentes e sequência semelhantes aos da Tabela 19.1. Essa seção geralmente inicia com uma descrição concisa da abordagem, às vezes sendo seguida por um diagrama esquemático ou por uma tabela para orientar o leitor (Tabela 19.2). Essa descrição deve explicitar de forma clara o delineamento do estudo e oferecer uma breve descrição dos participantes, das principais aferições, das intervenções, da duração do seguimento e dos desfechos principais.

TABELA 19.2 Exemplo de um cronograma de estudo para um ensaio randomizado sobre os efeitos da administração de testosterona nos fatores de risco de doenças cardíacas, câncer de próstata e fraturas

	VISITA DE RASTREAMENTO	RANDOMIZAÇÃO	3 MESES	6 MESES	12 MESES
História médica	X	–	–	–	X
Pressão arterial	X	X	X	X	X
Exame da próstata	X	–	–	–	X
Antígeno específico da próstata (PSA)	X	–	–	–	X
Lipídeos séricos	–	X	X	X	X
Marcadores inflamatórios	–	X	–	–	X
Densidade óssea	–	X	–	–	X
Marcadores de turnover ósseo	–	X	X	–	X
Força de preensão	–	X	X	X	X
Eventos adversos	–	–	X	X	X

A seção sobre a abordagem inclui uma breve justificativa para a pesquisa, apoiada por **dados preliminares** – estudos prévios realizados pelo investigador e por sua equipe que sugerem que o estudo proposto será bem sucedido. Deve-se enfatizar a importância dos trabalhos anteriores do grupo no assunto proposto e as razões pelas quais esses estudos merecem continuidade ou prorrogação. Resultados de estudos-piloto que sustentam a importância da questão de pesquisa e a **factibilidade** do estudo são importantes para vários tipos de propostas, especialmente quando a equipe de pesquisa possui pouca experiência anterior com os métodos propostos, quando a questão é inédita e quando há dúvidas quanto à factibilidade dos procedimentos propostos ou do recrutamento de participantes. Essa é uma oportunidade para mostrar que o investigador e a sua equipe têm a experiência e a expertise necessárias para conduzir o estudo.

Outros componentes específicos da seção sobre abordagem já foram discutidos. A seção sobre os **sujeitos do estudo** (Capítulo 3) deve definir claramente e fornecer uma justificativa para os critérios de inclusão e de exclusão e especificar o método de amostragem. É importante descrever como os participantes do estudo serão recrutados e assegurar aos revisores que os investigadores são capazes de arrolar o número desejado de participantes. Devem-se também descrever os planos para otimizar a adesão à intervenção (se aplicável) e às visitas do estudo.

A seção de abordagem deve incluir uma descrição de **procedimentos** importantes do estudo, como randomização e cegamento. A parte referente às **aferições** (Capítulo 4) deve descrever claramente como as variáveis preditoras, de desfecho e de confundimento serão medidas e em que ponto no estudo essas aferições serão feitas. Outros aspectos importantes são como as intervenções serão aplicadas e como o desfecho principal será avaliado e medido.

A seção de **estatística** geralmente começa com os **planos para análise**, que costumam ser organizados por objetivo específico. Ela pode ser elaborada seguindo uma sequência lógica; por exemplo, primeiro as tabulações descritivas e, então, as abordagens para a análise de associações entre variáveis. Após vem a uma discussão sobre **tamanho de amostra e poder estatístico** (Capítulos 5 e 6), que deve começar com a hipótese nula para o objetivo que irá determinar o tamanho de amostra do

estudo. As estimativas do tamanho de amostra e do poder estatístico se baseiam em pressupostos sobre a magnitude das associações que provavelmente serão detectadas e sobre a precisão das medições que serão feitas. Esses pressupostos devem ser justificados com base na literatura publicada ou em trabalhos preliminares que apoiam esses julgamentos. Pode ser útil incluir uma tabela ou figura que mostra como variações no tamanho de amostra, no poder estatístico ou em outros pressupostos influenciam o tamanho de amostra, para mostrar que o investigador fez escolhas adequadas. Os painéis de revisores do NIH costumam dar importância considerável à seção de estatística. Assim, é aconselhável ter um estatístico engajado na elaboração desse elemento da proposta.

Recomenda-se incluir uma tabela que **lista as visitas** do estudo ou contatos com o participante, o cronograma das visitas e que procedimentos ou aferições irão ocorrer em cada visita. Tal tabela fornece um panorama geral de todas as atividades do estudo (Tabela 19.2). As descrições do **gerenciamento de dados** e do **controle de qualidade** (Capítulos 16 e 17) devem abordar como os dados do estudo serão coletados, armazenados e editados, juntamente com planos para maximizar a qualidade dos dados e sua segurança.

A proposta deve fornecer um plano de trabalho realista e um **cronograma** indicando as datas em que cada fase importante do estudo será iniciada e finalizada (Figura 19.1). Cronogramas similares podem ser elaborados para as atividades da equipe e para outros componentes do projeto. Para estudos maiores, um **organograma** descrevendo a equipe de pesquisa pode indicar os níveis de autoridade e de responsabilidade, as linhas de hierarquia e também mostrar como será a dinâmica de trabalho da equipe.

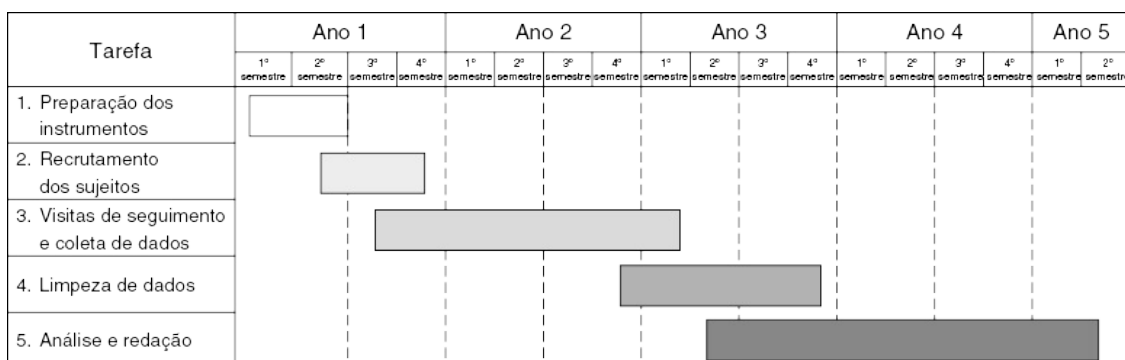



FIGURA 19.1 Um cronograma hipotético.

Embora essa não seja uma seção obrigatória, pode ser útil incluir uma discussão sobre as **limitações** do estudo proposto e as **abordagens alternativas**. Ao invés de simplesmente ignorar as potenciais falhas do estudo, o investigador pode decidir abordá-las explicitamente, discutindo as vantagens e desvantagens das várias decisões que precisaram ser tomadas para se chegar ao plano final. Apontar os desafios importantes e as potenciais soluções pode transformar eventuais críticas à proposta em pontos fortes. Entretanto, é um erro hiperenfatizar esses problemas, pois isso pode levar o revisor a focar de forma desproporcional nos aspectos mais fracos da proposta. O objetivo é tranquilizar o revisor de que o investigador antecipou todos os potenciais problemas e que tem uma abordagem realista e bem pensada para lidar com eles.

Componentes finais de uma proposta de grande porte

A seção sobre os **participantes** é dedicada às questões éticas levantadas pelo estudo, abordando temas como segurança, privacidade e confidencialidade. Essa seção contempla a maneira como os potenciais participantes da pesquisa serão informados sobre riscos e benefícios da participação e como será obtido seu consentimento para que possam tomar parte da pesquisa (Capítulo 14). Também descreve a inclusão de mulheres, crianças e participantes de grupos minoritários, como exigido para as propostas financiadas pelo NIH, e justifica a exclusão de qualquer desses grupos.

As **referências** indicam a familiaridade do investigador com a área de estudo. Elas devem ser abrangentes e ao mesmo tempo parcimoniosas e atualizadas – não uma lista exaustiva e não selecionada. Cada referência deve ser citada corretamente; erros nas citações ou má interpretação dos trabalhos desagradam os revisores familiarizados com a área.

Para alguns tipos de propostas, os **apêndices** podem ser úteis para incluir materiais técnicos e de apoio detalhados que foram mencionados apenas brevemente no texto. (Entretanto, para evitar que os apêndices sejam um meio de burlar o limite máximo de páginas, o NIH estabelece restrições claras para o seu uso.) Os apêndices podem incluir instrumentos de coleta de dados (como questionários) e protocolos clínicos, bem como

até três manuscritos e resumos que foram aceitos para publicação, mas ainda não publicados. Os revisores primários e secundários são os únicos membros do comitê de revisão que irão receber os apêndices. Portanto, todos os elementos importantes devem ser sumarizados sucintamente na proposta principal.

A finalidade e a importância de cada **consultor** devem ser descritas e acompanhadas de uma carta individual de aceite de cada um e de uma cópia de seu *curriculum vitae* resumido (*biosketch)*. (Os investigadores que serão pagos pelo auxílio não necessitam apresentar cartas, pois eles oficialmente já fazem parte da proposta.) Outras **cartas de apoio** devem também ser incluídas, como aquelas de pessoas que fornecerão acesso a equipamentos ou recursos. Deve ser incluída, também, uma explanação sobre os acordos programáticos e administrativos entre a instituição do investigador e as **instituições ou laboratórios colaboradores**, acompanhada de cartas de compromisso endereçadas ao investigador.

CARACTERÍSTICAS DE UMA BOA PROPOSTA

Uma boa proposta para financiamento deve apresentar vários atributos. O primeiro é a **qualidade científica** da estratégia de pesquisa: se parte de uma boa questão de pesquisa; utiliza delineamento e abordagens rigorosas e factíveis; e apresenta equipe de pesquisa experiente, apta e comprometida com o estudo. O segundo é a **clareza na apresentação**; uma proposta concisa e que prende a atenção do revisor, bem-organizada, cuidadosamente escrita, apresentada de maneira atraente e desprovida de erros leva o leitor a acreditar que a pesquisa será conduzida com um padrão semelhante de alta qualidade.

Os revisores estão em geral sobrecarregados com pilhas de propostas. Assim, o mérito do projeto deve se sobressair de maneira a ser compreendido com uma leitura rápida e superficial. Um **resumo** claro que segue os objetivos específicos, seções curtas com **subcabeçalhos** esclarecedores e o uso de **tabelas e figuras** para quebrar seções longas de texto fazem com que o revisor entenda os aspectos importantes da proposta. As diretrizes atuais do NIH sugerem iniciar os parágrafos com uma **frase-tópico** em negrito que torna claro o tema central, facilitando que revisores sobrecarregados compreendam os elementos essenciais da

proposta, examinando rapidamente as frases-tópico. É importante considerar os diversos pontos de vista e expertises dos revisores, incluindo detalhamento suficiente para convencer um revisor especialista sobre a relevância e sofisticação da proposta, ao mesmo tempo em que mantêm interessados um número maior de revisores que não têm conhecimento detalhado da área de investigação.

A maioria dos revisores reage negativamente a propostas exageradas e insensíveis à realidade. Propostas que valorizem demasiadamente a importância do projeto ou seu potencial impacto geram um certo ceticismo. O segredo está em escrever com entusiasmo procurando sempre ser realista quanto às limitações do projeto. Muitos revisores identificam com facilidade potenciais problemas no delineamento e na factibilidade de um projeto de pesquisa.

Uma rodada final de revisão científica por cientistas experientes que não estiveram envolvidos no desenvolvimento da proposta, em um momento em que ainda sejam possíveis modificações substanciais, pode ser de grande utilidade para a proposta, além de constituir uma experiência acadêmica compensadora. Também se recomenda que a proposta passe pelas mãos de alguém com boa capacidade de redação, para suplementar os programas de correção gramatical e ortográfica do editor de texto com sugestões sobre estilo e clareza.

BUSCANDO FORMAS DE APOIO À PESQUISA

É importante ficar alerta para opções de pesquisa que prescindam de uma proposta formal de financiamento. Por exemplo, um pesquisador iniciante poderá analisar pessoalmente bases de dados previamente coletados ou solicitar tempo de um cientista experiente de seu departamento para conduzir pequenos estudos. Conduzir uma pesquisa sem financiamento formal pode ser mais ágil e mais simples; no entanto, tem suas desvantagens, já que o projeto deverá ser de escopo limitado. Além disso, as instituições acadêmicas costumam basear suas decisões sobre avanços na carreira, em parte, na capacidade comprovada do investigador de angariar recursos para pesquisa. As quatro principais categorias de fontes de financiamento à pesquisa médica são:

- **Governo** (nos Estados Unidos, principalmente o NIH, mas também o Department of Veterans Affaris, o Centers for Disease Control and Prevention [CDC], a Agency for Healthcare Research and Quality [AHRQ], o Patient Centered Outcomes Research Institute [PCORI], o Department of Defense [DOD] e outras agências federais, estaduais ou municipais),
- **Fundações, sociedades de especialistas**, como o American Heart Association e o American Cancer Society, e **doadores individuais**,
- **Corporações privadas que visam ao lucro** (em geral as indústrias farmacêuticas e de equipamentos médicos) e
- **Recursos na própria instituição** (p. ex., oriundos da própria universidade do investigador).

Conseguir apoio de uma dessas fontes envolve um processo complexo e competitivo que favorece investigadores com **experiência** e **persistência**, de modo que os pesquisadores iniciantes devem procurar apoio de um mentor com essas características. Nas seções a seguir, focaremos nas principais dessas fontes.

Auxílios e contratos de pesquisa oferecidos pelo NIH

O NIH financia vários tipos de propostas de pesquisa. Os *R awards* (*R*01 e os auxílios menores *R*03 e *R*21) financiam projetos elaborados pelo investigador a partir de um tópico de pesquisa de sua escolha ou em resposta a um edital publicado por um dos institutos que compõem o NIH (ver www.nimh.nih.gov/research-funding/grants/research-grants-r.shtml). Os *K awards* (K23, K01, K08, K24 e os auxílios fornecidos localmente K12 e KL2) são um recurso excelente que permite pagar salário para treinamento e desenvolvimento de carreira de investigadores júniores, assim como um financiamento razoável para projetos de pesquisa (ver www.grants.nih.gov/training/careerdevelopmentawards.htm/).

Propostas desencadeadas por **iniciativas dos institutos** visam a estimular a pesquisa em áreas designadas pelos comitês consultivos do NIH e podem ser de dois tipos: Editais para Propostas (Requests for Proposals ou **RFPs**) ou Editais para Projetos Temáticos (Requests for

Applications ou **RFAs**). Em resposta a um RFP, o investigador firma contrato para executar certas atividades de pesquisa determinadas pelo NIH. Nos RFAs, o investigador submete projeto de pesquisa sobre um tema definido pelo NIH, mas ele próprio propõe a questão de pesquisa e o plano de estudo. Os RFPs usam o mecanismo de **contrato** para reembolsar gastos da instituição do investigador; e os RFAs usam o mecanismo de **auxílio** para apoiar as atividades de uma forma mais aberta.

Após a submissão de uma proposta, ela passa por um **processo de revisão** que inclui uma revisão administrativa inicial por funcionários do NIH, uma **revisão por pares** por um grupo de cientistas, recomendações sobre financiamento pelo conselho consultivo do instituto e decisão final sobre financiamento pelo diretor do instituto. As propostas de auxílio são geralmente avaliadas por uma de muitas "**seções de estudo**" (*study sections*) do NIH, que são grupos de revisores científicos com uma área de expertise específica, oriundos de diversas instituições de pesquisa norte-americanas. Cada uma dessas seções tem um foco específico e é composta por peritos na área de conhecimento, escolhidos de diferentes instituições de todo o país. Uma lista das seções de estudo e de seus atuais membros encontra-se disponível na página do NIH.

O processo do NIH para revisar e financiar propostas é descrito no endereço cms.csr.nih.gov. Quando um investigador submete uma proposta de auxílio ao NIH, ela é encaminhada pelo Centro para Revisão Científica (Center for Scientific Review ou CSR) a uma seção de estudo específica (Figura 19.2). As propostas são designadas a um revisor primário e dois ou mais revisores secundários; cada um deles atribui uma nota de 1 a 9 em relação à **relevância**, **caráter inovador**, **abordagem**, **investigadores** e **ambiente** e então uma nota global para o provável **impacto** do estudo. Um escore de "1" indica uma proposta excepcionalmente forte sem nenhum ponto fraco, e um escore de "9" indica uma proposta com deficiências graves substanciais e poucos pontos fortes. As notas atribuídas pelos revisores são reveladas à seção do estudo, e as propostas com escores na metade superior são discutidas com todo o comitê; as demais são "triadas" (não discutidas), sendo que algumas são postergadas para o próximo ciclo 4 meses após, dependendo do esclarecimento de alguns pontos que não haviam ficado claros. Após discussão, os revisores

novamente propõem notas (os escores podem ter mudado como resultado da discussão), e então todos os membros do comitê dão uma nota por meio de voto secreto. Dessas notas é feita uma média, que é multiplicada por 10, para produzir um escore global de 10 (melhor) a 90 (pior), que é utilizado por cada instituto para priorizar as decisões de financiamento.

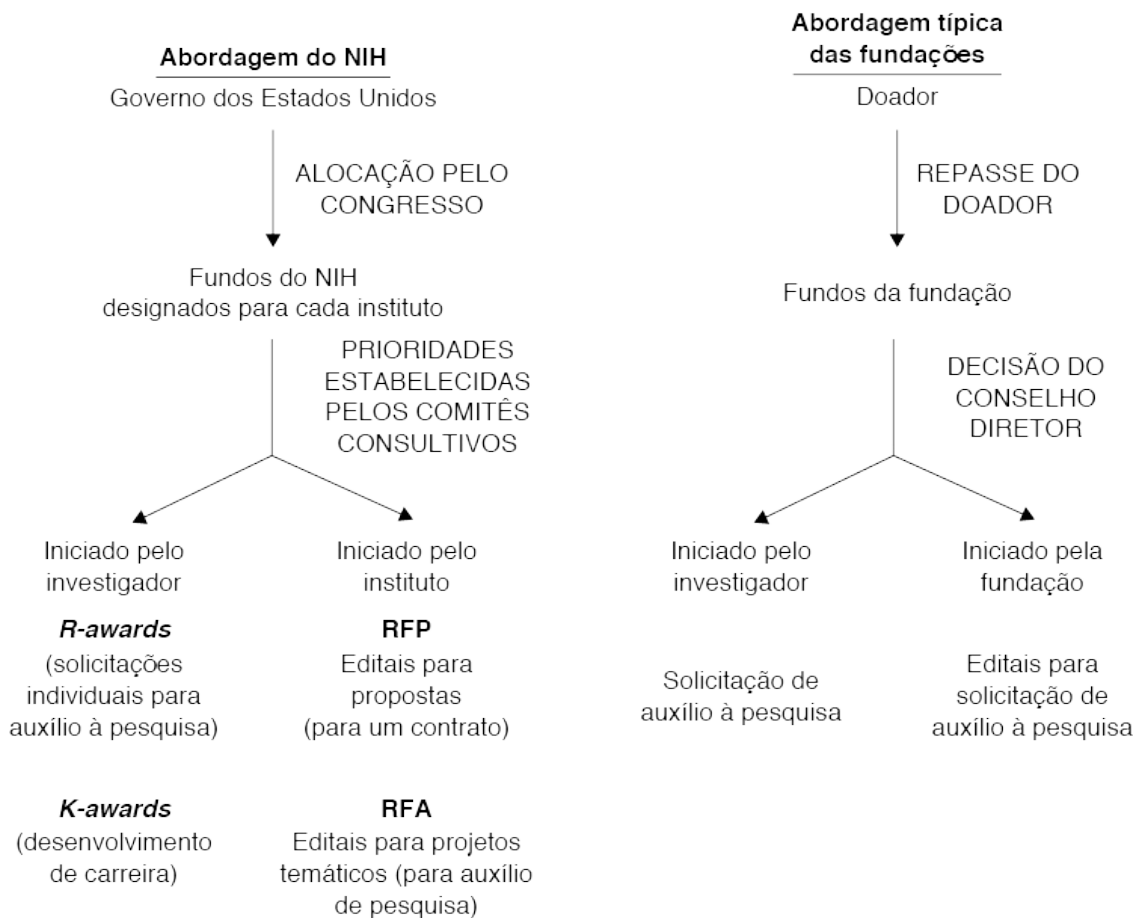



FIGURA 19.2 Esquema, em linhas gerais, das fontes e mecanismos para financiamento do NIH e das fundações.

Deve-se decidir com antecedência, com a ajuda de colegas mais experientes, que seção de estudo deveria revisar a proposta. As seções de estudo variam muito entre si, não somente quanto à área, mas também quanto à expertise dos avaliadores e à qualidade das propostas que estão competindo pelo financiamento. Embora a decisão não esteja completamente sob o controle do investigador, ele poderá influenciá-la envolvendo o administrador científico da agência de fomento com a qual

ele está colaborando para direcionar a proposta para a seção mais adequada.

Além de designar cada proposta de auxílio a uma seção de estudo específica, o CSR também a designa a um determinado **instituto** (ou centro) do NIH. Cada instituto então financia as propostas de auxílio a ele designadas, seguindo a ordem do escore de prioridade, que pode ser ajustada de acordo com a revisão do conselho consultivo e às vezes alterada pelo instituto (Figura 19.3). As propostas de investigadores iniciantes que ainda não receberam financiamento de pesquisa do NIH são financiadas com escores e pontos de corte de percentis menos rigorosos do que aquelas de investigadores mais estabelecidos. Se uma proposta for de interesse de mais de um instituto, os institutos às vezes dividem a responsabilidade pelo financiamento.

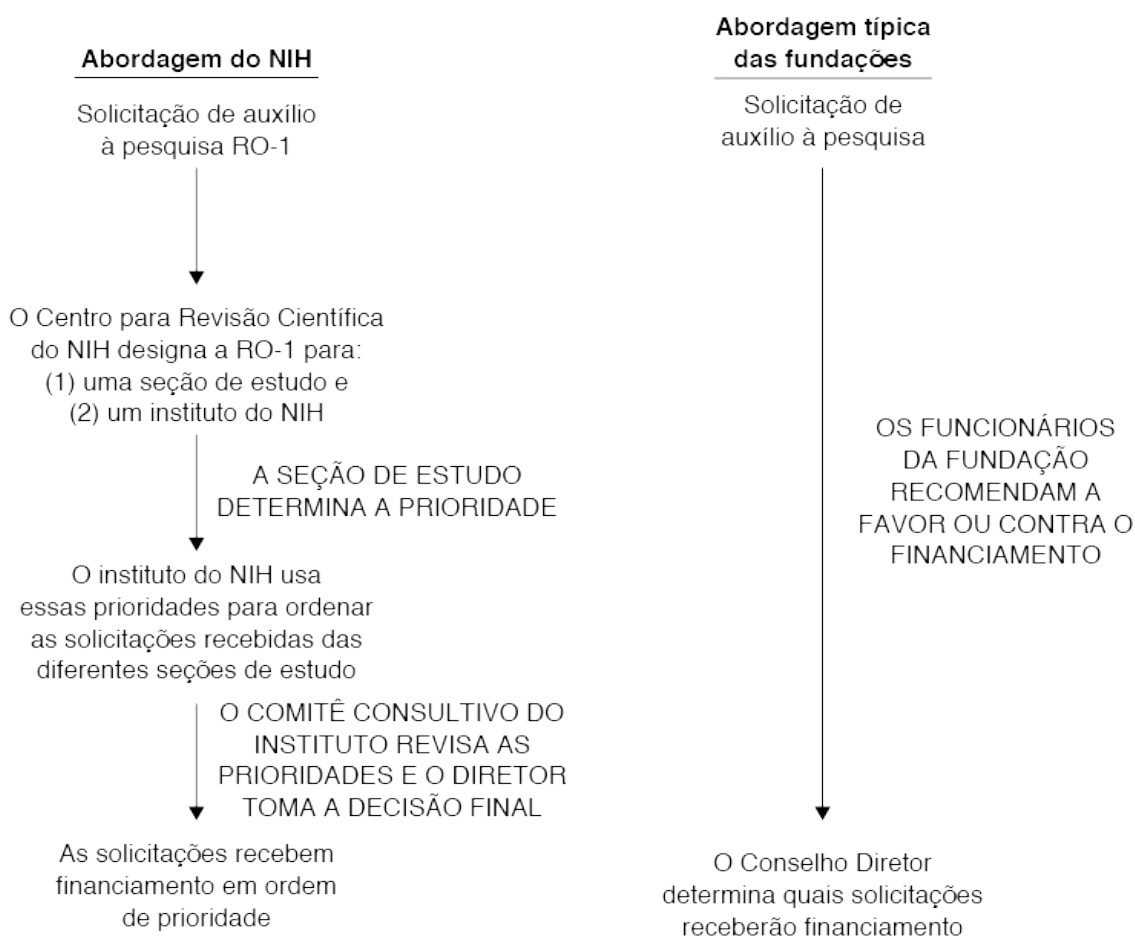



FIGURA 19.3 Procedimentos do NIH e das fundações para revisão das propostas

de auxílio à pesquisa.

Após ter sua solicitação avaliada, o investigador recebe uma notificação por escrito informando a decisão da seção de estudo. Esse **parecer descritivo** (*summary statement*) inclui o escore e comentários e críticas detalhados dos membros do comitê que revisaram a proposta.

As solicitações para o NIH que foram indeferidas, isto é, aquelas que não obtiveram financiamento, o que frequentemente acontece na primeira rodada, poderão ser revisadas e submetidas novamente apenas uma vez. Se as críticas e os escores dos revisores sugerirem que a proposta pode ser modificada de modo a se tornar mais atraente ao comitê, então uma nova versão, se reencaminhada, tem grande chance de obter financiamento. (Por outro lado, pode ser mais difícil aumentar o entusiasmo dos revisores se eles indicarem que a proposta é pouco inovadora ou relevante.) Os administradores científicos do instituto de interesse geralmente comparecem às reuniões da seção de estudo e é importante discutir a revisão com um deles logo após a reunião, pois os comentários escritos geralmente foram esboçados antes da reunião e podem não refletir questões que os membros da seção de estudo levantaram e que levaram a revisões nos escores.

Não é necessário que o investigador faça automaticamente todas as alterações indicadas pelos revisores; no entanto, ele deve fazer todo o possível para satisfazer as críticas dos revisores e justificar quando decidir não fazer alguma das mudanças sugeridas. O NIH limita a resposta às revisões a uma única página introdutória que descreve as alterações que foram feitas na proposta revisada. Um bom formato para essa página introdutória é sumarizar sucintamente as principais críticas do parecer descritivo em negrito ou itálico e então responder a elas com uma descrição concisa da alteração que foi feita na proposta. Para ajudar os revisores a focarem nessas revisões, as alterações devem ser marcadas, por exemplo, por uma linha vertical na margem esquerda do texto.

Verbas de fundações e sociedades de especialistas

As **fundações privadas** (como a The Robert Wood Johnson Foundation) geralmente restringem seus financiamentos a áreas específicas de interesse. Algumas fundações criadas para combater doenças específicas e

algumas **sociedades de especialistas** (como a American Heart Association e a American Cancer Society) também patrocinam programas de pesquisa, muitos deles designados a apoiar pesquisadores juniores. A quantia oferecida por essas fundações é bem inferior à do NIH, e muitas delas têm por objetivo financiar projetos meritórios que abordam tópicos ou utilizam metodologias que provavelmente não seriam financiados pelo NIH. Algumas fundações oferecem bolsas de desenvolvimento de carreira focadas em áreas específicas, como qualidade da atenção à saúde. O **Foundation Center** (http://fdncenter.org/) mantém um diretório de fundações, com informações de contato e sugestões para escrever propostas eficazes para essas fundações. As decisões sobre financiamento seguem procedimentos que variam de uma instituição para outra, mas geralmente a **resposta é rápida** para **propostas relativamente breves** (Figura 19.3). As decisões são tomadas por meio de um processo executivo, e não por processo de avaliação por pares. Os funcionários da fundação em geral produzem uma recomendação que é ratificada pelo conselho diretor.

Para determinar se uma fundação poderia ter interesse em uma proposta específica, o investigador deve pedir conselho a seus mentores ou então visitar a **página da fundação na internet**. A página na internet geralmente descreve os objetivos e propósitos da fundação e com frequência lista projetos recentemente financiados. Caso o investigador considere que uma fundação poderá ser uma fonte de apoio, o melhor a fazer é contatar o funcionário encarregado da fundação, descrever a ele o projeto, certificar-se de que há um potencial interesse e pedir orientação sobre como submeter a proposta. Muitas fundações pedem aos investigadores que enviem uma carta descrevendo o *background* e os principais objetivos do projeto, a qualificação dos investigadores e a duração e os custos aproximados da pesquisa. Caso a carta suscite interesse, a fundação poderá requisitar uma proposta mais detalhada.

Apoio da indústria

Empresas fabricantes de fármacos e equipamentos estão entre as principais fontes de recursos, especialmente para ensaios clínicos randomizados sobre novos tratamentos. Grandes empresas geralmente

aceitam propostas de pesquisa iniciadas pelo investigador, que podem incluir pequenos estudos sobre os efeitos ou mecanismos de ação de um tratamento ou estudos epidemiológicos sobre doenças de interesse à empresa. Muitas vezes elas fornecem a medicação ou o equipamento estudado, bem como um placebo idêntico, para um ensaio clínico proposto por um investigador que seja de interesse para a empresa. Elas podem também oferecer pequenos auxílios para financiar programas educativos em suas áreas de interesse. No entanto, de longe, a principal forma de apoio da indústria para a pesquisa clínica é por meio de contratos com investigadores principais de centros clínicos para arrolar participantes em **ensaios clínicos multicêntricos** que testam novos medicamentos e equipamentos. Esses ensaios clínicos de grande porte são às vezes delineados e gerenciados por um centro de coordenação acadêmico, mas costumam ser controlados pelo patrocinador corporativo, geralmente por meio de um contrato com uma organização de pesquisa clínica (*clinical research organization*, CRO).

Solicitações de auxílio para uma pesquisa ou programa educativo, ou para participar como centro de estudos em um ensaio clínico, geralmente iniciam pelo contato com o representante regional da indústria. Se a empresa estiver interessada no assunto, ela poderá solicitar que o investigador submeta uma proposta relativamente breve, proponha um orçamento e preencha outros formulários. A indústria costuma dar preferência a solicitações de "**formadores de opinião**", isto é, de clínicos ou de investigadores renomados, que estiveram envolvidos em pesquisas ou consultorias com a empresa, e cuja opinião possa influenciar a forma como outros clínicos prescrevem medicamentos ou utilizam equipamentos. Portanto, um jovem investigador que busca apoio da indústria deve geralmente procurar ajuda de algum mentor de renome na hora de contatar a indústria e submeter a proposta.

Os **contratos** para arrolar participantes em ensaios clínicos geralmente pagam aos investigadores principais do centro clínico um valor fixo para cada participante incluído no ensaio clínico multicêntrico, encerrando-se o arrolamento quando a meta global de número de sujeitos foi alcançada. Um investigador pode arrolar um número suficiente de participantes para receber financiamento que exceda seus gastos, e nesse caso ele pode ficar com o excedente para gastar da forma como ele achar melhor, mas

perderá dinheiro se recrutar um número muito pequeno de participantes para pagar a equipe e as despesas institucionais relacionadas ao estudo. Antes de decidir participar como centro em ensaios clínicos multicêntricos, o investigador deve ter a certeza de que o contrato será aprovado pelas instâncias administrativas e pelo CEP de sua instituição a tempo de poder incluir um número suficiente de participantes antes de o arrolamento ser encerrado.

O financiamento da indústria, especialmente de departamentos de *marketing*, é geralmente canalizado para assuntos e atividades que podem aumentar a venda dos seus produtos. Os achados dos ensaios clínicos gerenciados pela indústria são geralmente analisados pelos estatísticos da própria empresa e os manuscritos são muitas vezes escritos por redatores médicos contratados por ela.

Vários IPs dos centros são geralmente selecionados para serem coautores de publicações em revistas com revisão por pares. As regulamentações federais exigem que os autores tenham acesso aos dados (incluindo o direito de solicitar análises de dados do estudo inteiro), que eles tenham feito contribuições substanciais aos manuscritos e assumam a responsabilidade pelas conclusões. Encorajamos aos IPs que busquem a autoria para si e para seus coinvestigadores e, se bem sucedidos nessa tarefa, que cumpram essas **exigências de autoria**. Idealmente, os planos de análise, os manuscritos e as apresentações de estudos multicêntricos devem ser revisados e aprovados por um comitê de publicações que tenha diretrizes escritas e uma maioria de membros que não sejam empregados da empresa que está patrocinando o estudo.

Uma **vantagem** do apoio corporativo é que ele é a única maneira prática de se abordar algumas questões de pesquisa. Não haveria nenhuma outra fonte de financiamento disponível, por exemplo, para testar um novo antibiótico que ainda não foi lançado no mercado. Outra vantagem é a relativa rapidez para obtenção de financiamentos dessa fonte; as decisões sobre pequenas propostas iniciadas pelo investigador são geralmente tomadas dentro de poucos meses, e a indústria farmacêutica é geralmente ávida para integrar investigadores qualificados em ensaios clínicos multicêntricos. Os cientistas da empresa geralmente têm ampla experiência com a intervenção terapêutica em estudo e com metodologia de pesquisa que pode ser útil na hora de planejar as análises e interpretar

os resultados. Além disso, a maioria das companhias farmacêuticas procura manter uma reputação íntegra, que favorece suas relações com o FDA e sua imagem pública. A experiência de pesquisa, o apoio estatístico e os recursos financeiros que a empresa fornece podem aperfeiçoar a qualidade da pesquisa.

Apoio na instituição

As universidades geralmente possuem recursos próprios de pesquisa para seus investigadores. Esses auxílios geralmente oferecem um valor limitado, mas podem ser obtidos de maneira **muito mais rápida** (de semanas a meses) e com **maior chance de êxito** do que os auxílios oferecidos pelo NIH ou por fundações privadas. Os fundos institucionais podem se restringir a finalidades especiais, como estudos-piloto, que podem levar a financiamentos externos, ou à compra de equipamentos. Tais fundos são geralmente reservados a professores juniores e são uma oportunidade ímpar de um pesquisador iniciante adquirir experiência na condução de um projeto financiado.

RESUMO

1. A **proposta para auxílio** é uma versão expandida do plano detalhado de um estudo (**protocolo**) que é utilizada para solicitar financiamento e também informações orçamentárias, administrativas e de apoio requeridas pela agência de fomento.
2. Ao desenvolver uma proposta de pesquisa, o investigador deve procurar orientação junto a colegas seniores sobre a **questão de pesquisa** e sobre escolha de uma **agência financiadora**. Os passos seguintes são estudar cuidadosamente as **diretrizes** da agência e contatar um **administrador científico** desta em busca de orientações.
3. O processo de redigir uma proposta, geralmente muito mais demorado que o esperado, inclui a organização de uma **equipe** que detenha o conhecimento necessário; a designação de um **investigador principal (IP)**; a elaboração de um **roteiro da proposta** que atenda estritamente às **diretrizes da agência**; a confecção de um **cronograma** para as tarefas escritas, a obtenção de uma **proposta-modelo**; e a avaliação do processo por meio de **reuniões** regulares. A proposta deve ser

examinada por colegas experientes e conhecedores do assunto, revisada muitas vezes e **aprimorada** no final com atenção a detalhes.

4. Os elementos principais de uma proposta incluem o **resumo** (*abstract*), as **partes administrativas** centradas no orçamento, justificativa do orçamento, currículos resumidos (biosketches) e recursos, a parte extremamente importante dos **objetivos específicos**, bem como a **estratégia de pesquisa**, com suas seções sobre **relevância**, **aspectos inovadores** e **abordagem**, incluindo **pesquisas prévias** pelo próprio investigador.
5. Uma **proposta de qualidade** requer não só uma **boa questão de pesquisa**, um bom **plano de estudo** e uma boa **equipe de pesquisa**, mas também uma **apresentação clara**. A proposta deve ser clara e concisa, seguir uma sequência lógica e indicar as vantagens e desvantagens de **opções adotadas** no plano de estudo. Os **méritos** da proposta devem sobressair por meio da utilização de **subcabeçalhos**, **tabelas** e **diagramas**, para que não passem despercebidos por um revisor atarefado.
6. Há quatro fontes principais de apoio à pesquisa clínica:
   a. O **NIH** e outras fontes governamentais são as **principais fontes** de auxílio à pesquisa. Eles usam um sistema complexo de avaliação por pares e revisão administrativa que, apesar de moroso, financia uma ampla gama de **auxílios e contratos de pesquisa**, bem como **bolsas para desenvolvimento de carreira**.
   b. As **fundações** e as **sociedades de especialistas** muitas vezes demonstram interesse em questões de pesquisa promissoras que não obteriam financiamento por parte do NIH. Seus procedimentos de revisão são mais **rápidos**, porém mais **parciais** do que os do NIH.
   c. Os **fabricantes de medicamentos e equipamentos** são uma fonte importante de **recursos** para a pesquisa, geralmente canalizados para estudos conduzidos pela empresa sobre **novos fármacos** e **equipamentos médicos**. Porém, as corporações valorizam parcerias com cientistas renomados e apoiam algumas pesquisas iniciadas pelo investigador.
   d. Os **fundos institucionais fornecem pequenas quantias de dinheiro** com **maior rapidez** e são um excelente primeiro passo para **estudos-piloto** e a **investigadores iniciantes**.

Exercícios

Capítulo 1 Introdução: anatomia e fisiologia da pesquisa clínica

1. O Apêndice 1 fornece uma descrição do estudo da Fórmula Limitada Precoce ("FLP") realizado em dois centros clínicos acadêmicos na Califórnia com o objetivo de estimular a amamentação de recém-nascidos que tinham perdido ≥ 5% de seu peso corporal. Neste ensaio clínico randomizado, a proporção de mães que relataram amamentação exclusiva aos 3 meses a um entrevistador cego foi de 79% no grupo FLP em comparação com 42% no grupo-controle ($P = 0{,}02$) (Flaherman et al. *Pediatrics* 2013;131 [no prelo]). Para cada uma das afirmações a seguir, indique (1) se se trata de uma inferência de validade interna ou de validade externa; (2) se você considera que é uma inferência válida; e (3) quaisquer razões pelas quais ela pode *não* ser válida.
   a. Para as mulheres no estudo, o fornecimento da fórmula limitada precoce aumentou as taxas de amamentação em 3 meses.
   b. O fornecimento da fórmula limitada precoce a lactentes com $5% de perda de peso nas primeiras 36 horas nascidos em um hospital comunitário em Boston provavelmente levará a taxas maiores de amamentação aos 6 meses de idade.
   c. Com base nos resultados deste estudo, é provável que um esforço internacional para fornecer a fórmula à maioria dos recém-nascidos aumente o sucesso da amamentação e melhore a saúde dos recém-nascidos e suas mães.
2. Para cada um dos resumos a seguir, retirados de estudos publicados, escreva uma única frase que especifique o delineamento e a questão da pesquisa, incluindo o principal preditor, as variáveis de desfecho e a população da amostra.

a. Investigadores em Winston-Salem, Carolina do Norte, analisaram uma amostra aleatória de 2.228 estudantes de ensino médio com relação à frequência com que assistiam a programas de luta na televisão nas últimas duas semanas, e 6 meses mais tarde questionaram os mesmos estudantes sobre a ocorrência de brigas na escola e nos namoros. As chances ajustadas de o estudante referir ter brigado com a namorada ou com o namorado aumentaram em 14% para cada episódio de luta que ele relatou ter assistido nos 6 meses anteriores. (DuRant RH et al., *Pediatrics* 2006;118:e265-72.)
b. Para avaliar se a duração da amamentação protege as mulheres contra câncer de ovário, os investigadores analisaram 493 mulheres chinesas com câncer de ovário recém-diagnosticado e 472 outras mulheres hospitalizadas, todas as quais tinham amamentado pelo menos um filho. Eles encontraram uma relação de dose-resposta entre os meses totais de amamentação e risco reduzido de câncer de ovário. Por exemplo, as mulheres que haviam amamentado por pelo menos 31 meses tinham uma razão de chances de 0,09 (IC 95% 0,04, 0,19) em comparação com mulheres que amamentaram por menos de 10 meses (Sue et al. *Am J Clin Nutr* 2013; 97:354-359).
c. Para descobrir se uma associação entre a ingesta de gordura saturada na dieta e uma concentração reduzida de esperma em homens inférteis se estendia para a população geral, investigadores dinamarqueses coletaram amostras de sêmen e realizaram questionários sobre a frequência alimentar de homens jovens que consentiram no momento de seu exame para o serviço militar. Eles encontraram uma relação de dose-resposta significativa entre a ingesta autorrelatada de gordura saturada na dieta e concentrações reduzidas de esperma (p. ex., concentração de esperma 41% [IC 95% 4%, 64%] mais baixa no quartil mais alto de ingesta de gordura saturada em comparação com o mais baixo) (Jensen et al. *Am J Clin Nutr* 2013;97:411-418).
d. Não há tratamento medicamentoso sabidamente efetivo para os cerca de 20% de pacientes com diarreia por *Clostridium difficile* que tiveram recaída após tratamento com antibióticos. Investigadores em Amsterdã estudaram pacientes ≥ 18 anos de idade que apresentaram uma recaída de diarreia por *C. difficile* após

pelo menos um curso de terapia antibiótica adequada. Eles foram aleatoriamente designados (sem cegamento) para um de três regimes: um curso de 5 dias de vancomicina seguida por lavagem intestinal e infusão de uma suspensão de fezes de doadores voluntários por meio de sonda nasoduodenal ou um curso-padrão de 14 dias de vancomicina com ou sem lavagem intestinal no dia 4 ou 5. O estudo foi interrompido precocemente após uma análise interina ter mostrado que a taxa de cura sem recaída em 10 semanas era de 13 em 16 (81%) no grupo de fezes de doador, em comparação com 4 de 13 com vancomicina isoladamente e 3 de 13 com vancomicina mais lavagem (P , 0,001 para ambas as comparações) (van Nood et al. *N Engl J Med* 2013;368:407-415).

Capítulo 2 Elaborando a questão de pesquisa e desenvolvendo o plano de estudo

1. Considere a questão de pesquisa: "Qual a relação entre depressão e saúde?". Primeiro, transforme isso em uma descrição mais informativa que especifique delineamento de estudo, preditor, desfecho e população. Então, discuta se a questão e o delineamento escolhidos atendem os critérios FINER (Factível, Interessante, Nova (Inovadora), Ética, Relevante). Reescreva a questão e o delineamento para resolver quaisquer problemas relacionados a esses critérios.
2. Considere a questão de pesquisa: "O paracetamol causa asma?". Volte ao ano de 2000, quando apenas começava a fazer essa pergunta, e forneça descrições de uma frase para dois estudos observacionais e um ensaio clínico para abordar progressivamente essa questão de pesquisa. Certifique-se de que cada frase especifique delineamento do estudo, preditor, desfecho e população. Depois disso, para cada um deles, considere se a questão de pesquisa e o delineamento escolhidos preenchem os critérios FINER (Factível, Interessante, Nova (Inovadora), Ética, Relevante).
3. Use as ideias deste capítulo e seus próprios interesses para elaborar uma questão de pesquisa e conceber uma descrição de uma página para um estudo que você possa realizar. Ele preenche os critérios FINER? Discuta diferentes delineamentos, amostras populacionais e

variáveis com um colega, buscando otimizar a natureza FINER de seu estudo.

Capítulo 3 Escolhendo os sujeitos do estudo:especificação, amostragem e recrutamento

1. Um investigador está interessado na seguinte questão de pesquisa: “Quais fatores levam pessoas a começar a fumar?”. Ele decide-se por uma amostra transversal de estudantes do ensino médio. Para tanto, convida todos os alunos do 3º ano do ensino médio de uma escola de bairro e inclui no estudo aqueles que se apresentam como voluntários.
   a. Comente a adequação dessa amostra para a população-alvo de interesse.
   b. Suponha que o investigador tenha decidido evitar o viés associado com a escolha de voluntários delineando uma amostra aleatória de 25% de todo o 3º ano, e que 70% de toda a amostra tenha sido do sexo feminino. Se o número de alunos e alunas matriculados nessa escola for semelhante, a desproporção na distribuição do sexo representa um erro de amostragem. Isso poderia ter ocorrido por erro aleatório, erro sistemático, ou ambos? Justifique sua resposta.
2. Um investigador considera delineamentos para analisar o público de *shows* de *rock* a fim de determinar suas atitudes em relação ao uso de plugues auriculares durante os *shows* para proteger sua audição. Dê nome aos seguintes métodos de amostragem para a seleção de indivíduos que irão preencher um questionário, comentando sobre a factibilidade e a capacidade de generalização dos resultados para todas as pessoas que frequentam *shows* de *rock*.
   a. Para cada espectador que entrou no auditório, você pediu que jogasse um dado virtual (no telefone celular do investigador), selecionando todos aqueles que obtiveram o número 6.
   b. Para cada espectador que entrou no auditório, você pediu que jogasse um dado virtual, selecionando os homens que obtiveram 1 e as mulheres que obtiveram um número par.
   c. Os ingressos para o *show* foram numerados de forma seriada. Você selecionou todos os espectadores cujo ingresso terminou em 1.

d. Quando todos os espectadores já estavam sentados, você escolheu cinco fileiras aleatoriamente, por meio de um sorteio com um baralho de cartas, onde cada carta correspondia a uma das fileiras. Você selecionou todos os espectadores sentados nessas cinco fileiras.
e. Você entrevistou os primeiros 100 espectadores que entraram no auditório.
f. Alguns ingressos foram vendidos pelo correio e outros na bilheteria logo antes do *show*. Sempre que havia cinco pessoas ou mais esperando na fila da bilheteria para comprar ingressos, você selecionava a última pessoa da fila (a que tinha mais tempo disponível).
g. Quando os espectadores começaram a deixar o *show*, você selecionou aqueles que estavam dispostos e em condições de parar e responder às perguntas.

3. Edwards et al. (Edwards et al. *N Engl J Med* 2013;368:633-643) relataram a carga de infecção causada pelo metapneumovírus humano (MPVH) em crianças com menos de 5 anos. Os sujeitos eram crianças nos arredores de Cincinnati, Nashville e Rochester, NY, durantes os meses de novembro a maio, de 2003 a 2009, que buscaram atenção médica por doença respiratória aguda ou febre. Os pacientes internados que consentiram entraram no estudo de domingo a quinta-feira, os pacientes ambulatoriais, 1 ou 2 dias por semana, e os pacientes do setor de emergência, 1 a 4 dias por semana. Os autores combinaram a proporção de crianças com teste positivo em cada um dos locais com dados nacionais (do National Ambulatory Medical Care Survey e do National Hospital Ambulatory Care Survey) sobre a frequência de consultas por doença respiratória aguda ou febre na população para estimar a carga global de MPVH nos Estados Unidos. Eles estimaram que o MPVH foi responsável por 55 consultas e 13 atendimentos de emergência para cada 1.000 crianças anualmente.
a. Qual é a população-alvo do estudo?
b. Qual foi a população acessível e qual a sua adequação para que se faça uma generalização para uma população-alvo?
c. Qual foi o esquema de amostragem e qual a sua adequação para que se faça uma generalização para a população acessível?

d. Descreva em termos gerais de que forma o esquema de amostragem precisaria ser levado em conta ao calcular intervalos de confiança para as taxas de MPVH calculadas?

Capítulo 4 Planejando as aferições: precisão,acurácia e validade

1. Classifique as variáveis a seguir como dicotômica, nominal, ordinal, contínua ou discreta ordenada. Alguma delas poderia ser modificada para aumentar o poder estatístico? Como?
   a. História de infarto (presente/ausente)
   b. Idade
   c. Escolaridade (ensino médio completo/menos do que ensino médio)
   d. Escolaridade (número total de anos cursados)
   e. Raça
   f. Número de doses diárias de bebida alcoólica
   g. Depressão (nenhuma, leve, moderada, grave)
   h. Oclusão das artérias coronárias (%)
   i. Cor do cabelo
   j. Obeso(IMC $\geq$ 30) /não obeso (IMC < 30)
2. Um investigador está interessado na seguinte questão de pesquisa: “A ingesta de suco de fruta com 6 meses de idade prediz o peso corporal com 1 ano de idade?” Ele planeja um estudo de coorte prospectivo, medindo o peso corporal com uma balança infantil. Ele percebe os problemas listados a seguir na hora de pré-testar suas medições. Esses problemas são devidos à falta de acurácia, precisão ou ambos? Eles se devem principalmente à variabilidade do observador, do sujeito ou do instrumento? E o que pode ser feito quanto a isso?
   a. Durante a calibração da balança, um peso de referência de 10 kg acaba pesando 10,2 kg.
   b. A balança fornece resultados variáveis, mas ao pesar um peso de referência de 10 kg 20 vezes, o peso médio é 10,01 ± 0,2 kg (desvio-padrão).
   c. Alguns bebês ficam com medo e, quando tentam sair da balança, o observador os segura para poder completar a medição.
   d. Alguns bebês ficam muito agitados, fazendo com que o indicador

da balança sofra fortes oscilações.

e. Alguns bebês são levados para o exame logo após terem sido alimentados, e outros estão com fome; alguns dos bebês estão com as fraldas molhadas.

3. Um investigador está interessado em estudar o efeito da limitação das horas de trabalho de residentes de cirurgia. Uma área que ele quer abordar é o *burnout*, e ele planeja avaliá-la com duas questões (respondidas em uma escala de 7 pontos) retiradas de um questionário mais extenso: (a) "Com que frequência você se sente exausto pelo trabalho?" e (b) "Com que frequência você sente que ficou mais ríspido com as pessoas desde que iniciou sua residência?".

O investigador pretende avaliar a validade dessas questões para medir o *burnout*. Para cada uma das seguintes descrições, cite o tipo de validade que está sendo analisado:

a. Os residentes com maiores escores de exaustão tinham maior probabilidade de abandonar o programa no ano seguinte.

b. Estes itens parecem questões adequadas para avaliar o *burnout*.

c. Os escores de *burnout* aumentam durante os estágios mais árduos e diminuem durante as férias.

d. Um estudo prévio com mais de 10.000 estudantes de medicina, residentes e médicos formados mostrou que esses dois itens capturam quase completamente os domínios de *burnout* emocional e despersonalização de *burnout* conforme medidos pelo amplamente aceito (porém muito mais longo) Maslach Burnout Inventory (West et al. *J Gen Intern Med* 2009;24:1318-1321).

Capítulo 5 Preparando-se para estimar o tamanho de amostra:hipóteses e princípios básicos

1. Defina os conceitos destacados em azul.

Um investigador está interessado em delinear um estudo com um **tamanho de amostra** suficiente para determinar se há associação entre índice de massa corporal e câncer gástrico em mulheres de 50 a 75 anos de idade. Ele planeja um estudo de caso-controle com um número igual de casos e controles. A **hipótese nula** é de que não há diferença no

índice de massa corporal médio entre os casos de câncer gástrico e os controles; o investigador escolheu uma **hipótese alternativa** bilateral. Ele almeja um **poder estatístico** de 0,80 com um **nível de significância estatística** (α) de 0,05, para poder detectar uma **magnitude de efeito** de uma diferença no índice de massa corporal de 1 kg/m^2 entre os casos e os controles. A revisão da literatura mostra que a **variabilidade** do índice de massa corporal em mulheres é um desvio-padrão de 2,5 kg/m^2.

2. Analise as situações a seguir. Qual é um exemplo de um erro Tipo I? E de um erro Tipo II? E de nenhum?
   a. Um ensaio clínico randomizado mostra que os sujeitos tratados com uma nova medicação analgésica apresentam maior declínio nos escores de dor durante o estudo do que aqueles tratados com placebo ($P = 0{,}03$).
   b. Um estudo de 10 anos mostra que os 110 sujeitos fumantes não apresentaram maior incidência de câncer de pulmão do que os 294 sujeitos não fumantes ($P = 0{,}31$).
   c. Um investigador conclui que "o nosso estudo é o primeiro a demonstrar que o uso de álcool reduz o risco de diabetes em homens com idade inferior a 50 anos ($P < 0{,}05$)".

Capítulo 6 Estimando o tamanho de amostra e o poder estatístico: aplicações e exemplos

1. Retome o exercício 1 do Capítulo 5. Determine quantos casos de câncer gástrico seriam necessários para o estudo. E se os investigadores almejarem um poder estatístico de 0,90? Ou um nível de significância estatística de 0,01?

   Opcional: Imagine que o investigador somente tenha acesso a 60 casos. O que ele poderia fazer?

2. A força muscular diminui com a idade. Evidências preliminares sugerem que parte dessa perda de força muscular se deve à deficiência progressiva de deidroepiandrosterona (DHEA). Os investigadores planejam um ensaio clínico randomizado com a administração de DHEA ou placebo idêntico por 6 meses em indivíduos idosos, e com

medição subsequente da força muscular. Estudos prévios mostraram uma força de preensão média em idosos de 20 kg, com um desvio-padrão de 8 kg. Para um α (bilateral) = 0,05 e um β = 0,10, quantos sujeitos seriam necessários para demonstrar uma diferença de 10% ou mais na força dos grupos tratado e placebo? Quantos sujeitos seriam necessários para um β = 0,20?

3. No exercício 2, os cálculos de tamanho de amostra indicaram que seria necessário um número maior de sujeitos que o número possível de ser arrolado. Um colaborador lembra que os idosos apresentam variações na força de preensão, o que explicaria grande parte da variabilidade da força de preensão após o tratamento e poderia obscurecer o efeito terapêutico. Ele sugere medir a força na linha de base e novamente após o tratamento, usando a mudança na força como variável de desfecho. Um pequeno estudo-piloto mostra que o desvio-padrão da mudança na força durante um período de 6 meses é de apenas 2 kg. Quantos sujeitos seriam necessários por grupo usando esse delineamento, para α (bilateral) = 0,05 e β = 0,10?
4. Um investigador suspeita que alunos da terceira série disléxicos sejam com maior frequência canhotos do que os não disléxicos. Estudos anteriores mostraram que cerca de 10% da população é de canhotos e que a dislexia é uma condição rara. Planeja-se um estudo de caso-controle que irá selecionar todos os estudantes disléxicos em um distrito escolar como casos, com um número igual de estudantes não disléxicos selecionados aleatoriamente como controles. Qual tamanho de amostra seria necessário para mostrar que a razão de chances para dislexia é de 2,0 em estudantes canhotos comparados com estudantes destros? Pressuponha um α = 0,05 (bilateral) e um β = 0,20.
5. Um investigador procura determinar a média de QI de estudantes de medicina na sua instituição, com um IC de 99% de ± 3 pontos. Um pequeno estudo-piloto sugere que os escores de QI entre estudantes de medicina variam de aproximadamente 110 a 150. Qual o tamanho de amostra aproximado necessário?

Capítulo 7 Delineando estudos transversais e de coorte

1. Considere a questão de pesquisa “A deficiência moderada de vitamina B_{12} causa fraturas de bacia em idosos?”
   a. Faça um esboço de um plano de estudo para tratar dessa questão de pesquisa com um estudo de coorte prospectivo.
   b. Outra estratégia seria comparar os níveis de vitamina B_{12} em uma população clínica geriátrica de mulheres com fraturas de bacia no passado com os níveis em mulheres sem essas fraturas prévias. Comparando com essa abordagem transversal, liste pelo menos uma vantagem e uma desvantagem de seu estudo de coorte prospectivo.
   c. Seu estudo de coorte poderia ser delineado como um estudo retrospectivo? Como isso afetaria as vantagens ou desvantagens recém-listadas?
2. Sung et al. (Sung et al. *Am J Obstet Gynecol* 2009 May;200(5):557.e1-5) examinaram a associação basal entre a frequência de incontinência urinária e sintomas depressivos em 338 mulheres com sobrepeso ou obesas de pelo menos 30 anos de idade arroladas no ensaio clínico PRIDE (Program to Reduce Incontinence by Diet and Exercise). Eles relataram que as mulheres com sintomas depressivos (N = 101) relataram um número médio maior de episódios de incontinência por semana em comparação com as mulheres sem sintomas depressivos (28 *vs*. 23; $P = 0{,}005$).
   a. Que tipo de estudo é este?
   b. Uma possível explicação para isso é que a depressão aumenta a frequência da incontinência urinária. Quais são outras explicações para esta associação e de que forma alterações no delineamento do estudo ajudam a defini-las?

Capítulo 8 Delineando estudos de caso-controle

1. Para responder à questão de pesquisa “Qual a influência de uma história familiar de câncer de ovário no aumento do risco de desenvolver esse câncer?”, o investigador planeja um estudo de caso-controle.
   a. Como ele selecionaria os casos?
   b. Como ele selecionaria os controles?

c. Comente as potenciais fontes de viés na amostragem de casos e controles.
d. Como ele mediria a "história familiar de câncer de ovário" como variável preditora de interesse? Comente as possíveis fontes de viés nessa medida.
e. Que medida de associação e que teste de significância estatística ele usaria?
f. Você considera o método de caso-controle uma forma adequada para responder a questão de pesquisa? Discuta as vantagens e desvantagens do delineamento de caso-controle em relação às outras possibilidades para essa questão de pesquisa.

2. O investigador deseja pesquisar a relação entre jogar jogos de *videogame* envolvendo corridas de carro e o risco de estar envolvido em um acidente de carro real (como motorista).
   a. Suponha que a exposição de interesse seja os efeitos a longo prazo do uso habitual desses jogos. Como o investigador selecionaria os casos e controles e mediria a exposição para um estudo de caso-controle sobre essa questão?
   b. Agora imagine que a exposição de interesse é se o uso de tais jogos na hora imediatamente anterior ao ato de dirigir aumenta o risco a curto prazo. Qual o delineamento para estudos de efeitos a curto prazo de exposições intermitentes? Descreva de que forma esse estudo seria realizado para essa questão de pesquisa.

Capítulo 9 Fortalecendo a inferência causal em estudos observacionais

1. O investigador realiza um estudo de caso-controle para tentar responder à seguinte questão de pesquisa: "Um maior consumo de frutas, verduras e legumes diminui o risco de doença coronariana?". Imagine que seu estudo mostre que os indivíduos no grupo-controle relatam consumir mais desses alimentos do que os indivíduos com doença coronariana.

   Quais são as possíveis explicações para essa associação inversa entre o consumo de frutas, verduras e legumes e doença coronariana? Dê atenção especial à possibilidade de que a associação entre o consumo de

frutas, verduras e legumes e doença coronariana seja confundida pela atividade física (i.e., os indivíduos que comem mais frutas, verduras e legumes também fazem mais atividade física, e esse é o motivo para sua menor taxa de doença coronariana). Que estratégias poderiam ser usadas para lidar com a atividade física como possível confundidor, e quais as vantagens e desvantagens de cada uma dessas estratégias?

2. Um estudo da rede PROS (Pediatric Research in Office Settings) de pediatras concluiu que entre lactentes com menos de 3 meses trazidos ao pediatra por febre, os meninos não circuncisados tinham cerca de 10 vezes mais risco de infecção do trato urinário em comparação com meninos circuncisados (Newman et al.. *Arch Pediatr Adolesc Med* 2002 Jan;156(1):44-54), uma associação que tem sido vista em vários estudos. É interessante notar que os meninos não circuncisados no estudo pareciam apresentar um risco *menor* de infecções na orelha (razão de risco = 0,77; P = 0,08). Explique de que forma a inclusão apenas de bebês com febre no estudo poderia introduzir uma associação entre circuncisão e infecções na orelha a qual não está presente na população geral de lactentes.
3. No exercício 1 do Capítulo 2, pedimos que você sugerisse estudos para avaliar a questão sobre se o paracetamol causa asma. Um mecanismo proposto para tal associação é a depleção de glutationa induzida pelo paracetamol, a qual protege os pulmões de dano oxidativo que pode levar à inflamação. Descreva de maneira sucinta como você poderia tirar vantagem da variação em genótipos antioxidantes maternos para aumentar a inferência de que a associação entre uso materno de paracetamol e asma na prole é causal.

Capítulo 10 Delineando um ensaio clínico randomizado cego

1. Um extrato herbal, a huperzina, tem sido usado na China para tratamento da demência, e estudos preliminares em animais e humanos mostraram resultados promissores. O investigador gostaria de testar se esse novo tratamento poderia diminuir a progressão da doença de Alzheimer. Estudos anteriores mostraram que o nível plasmático de Abeta (1-40) é um biomarcador para a doença de Alzheimer: níveis

elevados estão associados a um risco significativamente maior de desenvolver demência, e os níveis de Abeta (1-40) aumentam com a progressão da demência. Ao planejar um ensaio clínico para testar a eficácia da huperzina na prevenção de demência em idosos com comprometimento cognitivo leve, o investigador considera duas potenciais medidas de desfecho: mudança nos níveis de Abeta (1-40) ou incidência de um diagnóstico clínico de demência.

a. Cite uma vantagem e uma desvantagem de se usar Abeta (1-40) como desfecho principal de seu ensaio clínico.

b. Cite uma vantagem e uma desvantagem de se usar o diagnóstico clínico de demência como desfecho principal do ensaio clínico.

2. Um grande ensaio clínico sobre a huperzina (> 200 pessoas por braço de estudo) está sendo planejado. O objetivo principal é testar se esse extrato herbal diminui a incidência de um diagnóstico clínico de demência em homens e mulheres idosos com comprometimento cognitivo leve.

a. Espera-se que a huperzina cause, em alguns pacientes, sintomas gastrintestinais, como diarreia, náusea e vômitos. Descreva um plano para avaliar os efeitos adversos desse novo tratamento sobre sintomas ou doenças que não sejam a demência.

b. Descreva um plano geral para coleta dos dados na linha de base: que tipos de informação deverão ser coletados?

c. Indivíduos portadores de um alelo Apoε4 têm um risco aumentado de demência. Cite um motivo a favor e um contra o uso de uma randomização em blocos e estratificada, em vez de uma randomização simples para garantir um balanço adequado de pessoas com o genótipo Apoε4 nos grupos de tratamento e controle.

Capítulo 11 Delineamentos alternativos para o ensaio clínico e tópicos relacionados à implementação

A finasterida tópica é moderadamente eficaz para tratar a calvície masculina e foi aprovada pelo FDA para esse fim. Demonstrou-se que as estatinas aumentam o crescimento capilar em roedores e que agem por um mecanismo diferente do que a finasterida. Imagine que uma nova empresa

deseja obter aprovação do FDA para a comercialização de uma nova estatina tópica (HairStat) para o tratamento da calvície masculina.

1. Descreva um ensaio clínico de fase I do HairStat para a calvície masculina. Qual(is) seria(m) o(s) grupo(s) de tratamento? Quais seriam os desfechos?
2. A empresa deseja comparar a eficácia do HairStat com aquela da finasterida. Cite pelo menos uma vantagem e uma desvantagem das seguintes estratégias para testar a eficácia relativa da finasterida e da estatina tópica.
   a. Randomizar homens calvos para finasterida ou estatina tópica.
   b. Em um delineamento fatorial, alocar aleatoriamente homens para (1) finasterida e HairStat, (2) finasterida e HairStat-placebo, (3) finasterida-placebo e HairStat, ou (4) placebo duplo.
3. Imagine que a empresa esteja planejando um ensaio clínico controlado por placebo com duração de 1 ano sobre o HairStat para tratamento da calvície. O desfecho é a mudança na determinação da quantidade de cabelo por meio de fotografias da região calva que recebe tratamento. Visitas de seguimento (com fotografias) são agendadas a cada 3 meses. Elabore um plano – contendo pelo menos dois elementos – para encorajar a adesão ao estudo e o retorno às visitas para avaliar o desfecho.
4. Vinte por cento dos homens no ensaio clínico sobre HairStat *versus* placebo não retornaram para a visita de seguimento de 3 meses, e 40% interromperam o tratamento antes de completar 1 ano. Em alguns casos, o motivo para isso foi o desenvolvimento de um *rash* no couro cabeludo. Cite uma desvantagem e uma vantagem de se analisar o efeito do tratamento sobre o crescimento capilar por meio de uma análise por intenção-de-tratar estrita.
5. Na análise por intenção-de-tratar, o HairStat aumentou o crescimento capilar (medido por avaliadores cegos com base em comparações de fotografias na linha de base e em um ano) em 20% (P = 0,06). Análises subsequentes mostraram que o crescimento capilar foi 45% maior com HairStat em homens com idade inferior a 40 anos do que em homens mais velhos (P = 0,01 naquele subgrupo). Quais são os problemas com a conclusão da empresa de que o HairStat é eficaz para o tratamento da calvície em homens com idade inferior a 40 anos?

Capítulo 12 Delineando estudos sobre testes médicos

1. Você está interessado em estudar a velocidade de sedimentação globular (VSG) como um teste para doença inflamatória pélvica (DIP) em mulheres com dor abdominal.
   a. Para isso, você precisará reunir grupos de mulheres com e sem DIP. Qual seria a melhor forma para amostrar essas mulheres?
   b. Que vieses poderiam surgir nos resultados se o diagnóstico final de DIP fosse usado como padrão-ouro e aqueles que fizeram o diagnóstico soubessem do resultado da VSG?
   c. Você descobre que a sensibilidade de uma VSG maior do que 20 é de 90%, mas a especificidade é de apenas 50%. Por outro lado, a sensibilidade de uma VSG maior do que 50 é de apenas 75%, mas a especificidade é de 85%. Que ponto de corte deveria ser usado para definir uma VSG anormal?
2. Você está interessado em estudar o rendimento diagnóstico de uma tomografia computadorizada (TC) do crânio em crianças que comparecem ao setor de emergência por traumatismos craniencefálicos. Você usa um banco de dados do serviço de radiologia para encontrar os registros de todas as TCs feitas em pacientes com menos de 18 anos solicitadas pelo setor de emergência para avaliar esse tipo de lesão. Você então revisa os registros do serviço de emergência de todos os indivíduos com TC anormal para determinar se a anormalidade poderia ter sido prevista no exame físico.
   a. De 200 TCs, 10 mostram lesões intracranianas. No entanto, você determina que, em 8 das 10 TCs, o exame neurológico do paciente havia apontado alterações focais ou do estado mental. Uma vez que apenas dois pacientes tiveram TCs anormais que não poderiam ter sido antecipadas pelo exame físico, você conclui que o rendimento diagnóstico de lesões intracranianas "inesperadas" foi de apenas 2 em 200 (1%). O que há de errado com essa conclusão?
   b. O que há de errado em usar lesões intracranianas identificadas pela TC como variável de desfecho para esse estudo sobre rendimento

diagnóstico?

c. Quais seriam algumas vantagens de se estudar os efeitos do teste sobre a tomada de decisões clínicas, em vez de apenas o rendimento diagnóstico?

3. Agora você gostaria de estudar a sensibilidade e a especificidade de alterações neurológicas focais para predizer lesões intracranianas. (Devido ao pequeno tamanho de amostra de pacientes com lesões intracranianas, você estende a amostra a pacientes atendidos em outros serviços de emergência.) Um problema que surge ao se estudar alterações neurológicas focais é que as crianças que apresentam essas alterações têm maior probabilidade de fazerem TCs do que aquelas que não apresentam essas alterações. Explique como e por que isso irá afetar a sensibilidade e a especificidade dos achados se:
   a. Apenas crianças que realizarem TC forem incluídas no estudo.
   b. Crianças elegíveis com traumatismos cranioencefálicos que não fizerem uma TC forem incluídas, pressupondo-se que não apresentavam lesão intracraniana caso se recuperassem sem intervenção neurocirúrgica.

Capítulo 13 Pesquisas com dados existentes

1. Considere a questão de pesquisa “Nos Estados Unidos, indivíduos de origem latino-americana têm taxas maiores de doença da vesícula biliar do que indivíduos brancos, afro-americanos ou asiáticos?”. Que bases de dados podem possibilitar que você determine as taxas de doença da vesícula biliar de acordo com raça, idade e sexo de forma rápida e econômica?
2. Um *fellow* de pesquisa ficou interessado em estudar se a disfunção renal leve ou moderada aumenta o risco de eventos coronarianos e morte. Devido aos custos e à dificuldade de se conduzir um estudo para gerar dados primários, ele buscou um banco de dados existente que continha as variáveis necessárias para responder à sua questão de pesquisa. Ele descobriu que o Cardiovascular Health Study (CHS), um estudo de coorte de grande porte, multicêntrico, financiado pelo NIH sobre preditores de doença cardiovascular em mulheres e homens idosos continha todas as variáveis necessárias para a análise dessa

questão. Seu mentor lhe apresentou investigadores importantes do CHS, que o ajudaram a preparar e a submeter uma proposta de análise que foi aprovada pelo Comitê Diretor do CHS.

a. Quais são as vantagens de tal abordagem para estudar essa questão de pesquisa?

b. Quais são as desvantagens?

3. Um investigador está interessado em estudar se os efeitos do tratamento na pós-menopausa com estrogênios ou com moduladores seletivos do receptor de estrogênio (SERMs) variam dependendo dos níveis estrogênicos endógenos. Como ele poderia responder a essa questão usando um estudo suplementar?

Capítulo 14 Abordando questões éticas

1. A questão de pesquisa é a identificação de genes associados a um maior risco de desenvolver diabetes tipo 2. O investigador descobre que amostras de sangue armazenadas e dados clínicos estão disponíveis em um grande estudo de coorte prospectivo sobre fatores de risco para doença coronariana. Esse estudo coletou dados basais sobre dieta, atividade física, características clínicas e medidas de colesterol e hemoglobina A1c. Dados de seguimento estão disponíveis sobre desfechos coronarianos e desenvolvimento de diabetes. O estudo proposto irá realizar sequenciamento de DNA em participantes; não serão necessárias novas amostras de sangue.

a. O estudo pode ser realizado usando apenas o consentimento informado original do estudo de coorte?

b. Se o consentimento original não incluiu a permissão para esse tipo de estudo, como os investigadores podem realizar o estudo planejado?

c. Ao delinear novos estudos que irão armazenar amostras de sangue, como os investigadores podem planejar estudos futuros que usam seus dados e amostras?

2. O investigador planeja um ensaio clínico randomizado de fase III sobre um novo quimioterápico promissor no tratamento do câncer de cólon. Para reduzir o tamanho da amostra, ele gostaria de realizar um estudo controlado por placebo, em vez de compará-lo com o

tratamento atual.

a. Quais são as preocupações éticas sobre um controle com placebo nessa situação?
b. É possível realizar um estudo controlado por placebo de forma eticamente aceitável?

3. O investigador planeja um estudo de viabilidade para um futuro ensaio clínico sobre uma vacina contra o HIV. Os objetivos são determinar (1) se é possível recrutar uma coorte de participantes com alta taxa de soroconversão, apesar de um aconselhamento moderno para a prevenção do HIV e (2) se o seguimento nessa coorte será suficientemente elevado para realizar o ensaio clínico da vacina. Os participantes serão pessoas com risco aumentado para o HIV, incluindo usuários de drogas injetáveis, pessoas que fazem sexo por dinheiro e outras pessoas com múltiplos parceiros sexuais. A maioria dos participantes será de indivíduos com baixa escolaridade e baixo nível de conhecimentos relacionados à saúde. O estudo será um estudo de coorte observacional, com seguimento de dois anos para determinar as taxas de soroconversão e acompanhamento.
   a. O que as normas federais exigem que seja informado aos participantes como parte do consentimento informado?
   b. Que passos devem ser tomados para garantir que o consentimento seja realmente informado nesse contexto?
   c. Qual é a responsabilidade dos investigadores durante esse estudo observacional para reduzir o risco de HIV nesses participantes de alto risco?

Capítulo 15 Elaborando questionários, entrevistas e pesquisas *on-line*

1. Como parte de um estudo sobre consumo de álcool e força muscular, um investigador planeja usar o seguinte item em um questionário de autopreenchimento para determinar o uso corrente de bebidas alcoólicas:

“Quantos *drinks* de cerveja, vinho ou destilados você toma a cada dia?”

○ 0
○ 1-2

○ 3-4
○ 5-6
○ 7-8

Descreva brevemente pelo menos dois problemas com este item.

2. Escreva algumas questões para um questionário de autopreenchimento que irá determinar de uma forma melhor o uso corrente de bebidas alcoólicas.
3. Comente as vantagens e desvantagens de um questionário de autopreenchimento *versus* uma entrevista estruturada para avaliar o comportamento sexual de risco.

Capítulo 16 Gerenciamento de dados

1. Este exercício refere-se aos primeiros seis itens do exemplo de questionário sobre tabagismo no Apêndice 15.1. Você tem respostas para três sujeitos do estudo:

ID do sujeito	Descrição da história de tabagismo
1001	Começou a fumar aos 17 anos e continuou fumando uma média de 30 cigarros/dia desde então
1002	Começou a fumar aos 21 anos e fumou 20 cigarros/ dia até parar, há três anos, aos 45 anos de idade
1003	Fumou alguns cigarros (< 100) quando estava no ensino médio

Elabore uma tabela de dados contendo as respostas desses sujeitos às primeiras seis questões do Apêndice 15.1. A tabela deve conter três linhas (uma para cada sujeito) e sete colunas (uma para a ID do sujeito e uma para cada uma das seis questões).

2. O estudo PHTSE (Pre-Hospital Treatment of Status Epilepticus) (Lowenstein et al. *Control Clin Trials* 2001;22:290-309; Alldredge et al. *N Engl J Med* 2001;345:631-637) foi um ensaio clínico randomizado cego sobre lorazepam, diazepam, ou placebo para o tratamento pré-hospitalar do estado de mal epiléptico. O desfecho principal foi o término das convulsões até a chegada ao hospital. Para arrolar os pacientes, os paramédicos contataram médicos do hospital de referência pelo rádio. A seguir estão os formulários para coleta de dados pelos médicos do hospital de referência sobre dois pacientes

arrolados:

PHTSE

Formulário para Coleta de Dados por Médicos do Hospital

ID do Sujeito no PHTSE:

Administração da Medicação: *189*

Kit da Medicação nº: *A322*

Data e Hora da Administração: *3 / 12 / 94* *17 : 39*
(Usar o formato de 24 horas)

Avaliação Durante o Transporte
Convulsão interrompida:
Hora da Interrupção: *17 : 44*
(Usar o formato de 24 horas)

Avaliação Final ("Final da Corrida")
Hora da Chegada ao Hospital *17 : 48*
(Usar o formato de 24 horas)

Ao chegar ao hospital:
[X] 1 Continuação da atividade convulsiva (convulsões tônico-clônicas ativas)
[ ] 0 Interrupção da atividade convulsiva (convulsões tônico-clônicas ativas)
- [ ] 1 Ausente
- [ ] 2 Sons ininteligíveis
- [ ] 3 Palavras inapropriadas
- [ ] 4 Fala confusa
- [ ] 5 Orientado

BIPAR O FUNCIONÁRIO DO ESTUDO PHTSE DE PLANTÃO!!

PHTSE

Formulário para Coleta de Dados por Médicos do Hospital

ID do Sujeito no PHTSE:

Administração da Medicação | 410

Kit da Medicação nº: | B536

Data e Hora da Administração *12 / 01 / 98* *01 : 35*
(Usar o formato de 24 horas)

Avaliação Durante o Transporte
[*X*]Convulsão interrompida:
Hora da Interrupção: *01 : 39*
(Usar o formato de 24 horas)

Avaliação Final ("Final da Corrida")
Hora da Chegada ao Hospital: *01 : 53*
(Usar o formato de 24 horas)

Ao chegar ao hospital:
[ ] 1 Continuação da atividade convulsiva (convulsões tônico-clônicas ativas)
[*X*] 0 Interrupção da atividade convulsiva (convulsões tônico-clônicas ativas)
Escala de Glasgow: Componente Verbal
[ ] 1 Ausente
[ ] 2 Sons ininteligíveis
[ ] 3 Palavras inapropriadas
[*X*] 4 Fala confusa
[ ] 5 Orientado

BIPAR O FUNCIONÁRIO DO ESTUDO PHTSE DE PLANTÃO!!

a. Organize os dados desses formulários em uma tabela com duas linhas.
b. Crie um dicionário de dados com nove campos para a tabela de dados recém-construída.
c. Os formulários de coleta de dados em papel foram preenchidos por médicos atarefados, chamados do serviço de emergência para uma sala de rádio. Quais são as vantagens e as desvantagens de se usar um formulário na tela do computador no lugar de um formulário em papel? Se você fosse responsável pelo delineamento do estudo, qual você usaria?

3. Os formulários de coleta de dados em papel no exercício 2 incluem uma questão sobre se a atividade epiléptica continuou após a chegada ao hospital (desfecho principal do estudo). Esse item recebeu o nome de campo *AtivConvChegHosp* e foi codificado como 1 para *sim* (continuação da atividade epiléptica) e 0 para *não* (interrupção da atividade epiléptica).

   Interprete os valores médios para *AtivConvChegHosp* como mostrado abaixo:

	AtivConvChegHosp (1 = Sim, convulsão continuou; 0 = Não, convulsão interrompida)	
	N	**Média**
Lorazepam	66	0,409
Diazepam	68	0,574
Placebo	71	0,789

Capítulo 17 Implementação do estudo e controle de qualidade

1. Um investigador estudou a questão de pesquisa “Quais são os preditores de óbito após a hospitalização para infarto do miocárdio?”. Pesquisadores assistentes coletaram dados detalhados a partir de prontuários e conduziram entrevistas abrangentes com 120 pacientes hospitalizados seguidos por um período de um ano. Cerca de 15% dos pacientes morreram ao longo do período de seguimento. Completada a coleta, um dos pesquisadores assistentes inseriu os dados em um computador usando um programa-padrão de planilhas eletrônicas. Terminada essa etapa, o investigador deu início à análise de dados. Para sua surpresa, descobriu que entre 10 e 20% dos dados sobre algumas variáveis preditoras estavam faltando e um número considerável parecia não fazer sentido. Somente 57% dos sujeitos da amostra haviam sido examinados durante o primeiro ano de seguimento, data agora vencida há mais de um ano para alguns sujeitos. Você é solicitado a fazer uma consultoria para esse projeto.
   a. No presente momento, o que o investigador poderá fazer para melhorar a qualidade de seus dados?

b. Descreva resumidamente pelo menos três maneiras para que ele possa reduzir os valores faltantes e os erros em seu próximo estudo.

Capítulo 18 Estudos comunitários e internacionais

1. Um investigador deseja estudar as características e o curso clínico de pacientes com dor abdominal de etiologia desconhecida. Ele planeja recrutar pacientes com dor abdominal cuja causa específica não foi identificada após uma bateria de exames. Há duas opções para o recrutamento dos sujeitos: 1) o ambulatório de gastrenterologia de seu centro médico universitário ou 2) uma rede local de unidades de atenção primária à saúde. Quais são as vantagens e as desvantagens de cada abordagem?
2. Um investigador foi designado para trabalhar em conjunto com o Ministério da Saúde chinês em um novo programa de prevenção de doenças relacionadas ao fumo na China. Das questões de pesquisa a seguir, quais delas exigem pesquisa local, não bastando pesquisas conduzidas em outros contextos?
   a. Qual é a prevalência e a distribuição do tabagismo na China?
   b. Quais são as doenças causadas pelo fumo?
   c. Quais são as estratégias mais eficientes para estimular as pessoas a pararem de fumar?

Capítulo 19 Redigindo uma proposta para solicitar financiamento de pesquisa

1. Faça uma busca na internet na página do NIH (*http://grants.nih.gov/grants/oer.htm*) para localizar pelo menos três tipos de *R-awards* iniciados pelo investigador.
2. Faça uma busca na internet por fundações que poderiam estar interessadas no tema de sua pesquisa. Cite pelo menos duas.
3. Contate mentores e colegas para encontrar um protocolo de pesquisa que avalie uma questão em sua área de interesse e que tenha sido financiado. Leia cuidadosamente este protocolo.

Respostas dos exercícios

Capítulo 1 Introdução: anatomia e fisiologia da pesquisa clínica

1a. Esta é uma inferência de validade interna (porque se refere às mulheres no estudo) e que provavelmente é válida. Porém, ela poderia ser inválida se outro fator além da fórmula limitada precoce (FLP) causasse a diferença nas taxas de amamentação (p. ex., se a intervenção-controle afetasse de forma adversa a amamentação), se a amamentação autorrelatada não refletisse a taxa real de amamentação ou se a associação não fosse causal (o P = 0,02 não descarta a ocorrência ao acaso).

1b. Esta é uma inferência de validade externa (porque envolve a generalização para fora do estudo) que pode ser válida. Porém, além das ameaças à validade interna antes descritas (que também podem ameaçar a validade externa), é provável que mulheres que dão à luz em hospitais da comunidade e em outras partes do país possam responder de maneira diferente à intervenção, ou que outros médicos que oferecem a FLP possam realizar a intervenção de modo diferente daquele feito no estudo original, ou os benefícios podem não durar 6 meses.

1c. Esta é uma inferência de validade externa que se estende muito além da população e da intervenção que foram estudadas, e provavelmente não é válida. Isso envolve a generalização não apenas para outras mães e recém-nascidos de outros locais, mas também inclui recém-nascidos que não perderam 5% de seu peso corporal; expande a intervenção além da fórmula limitada precoce para fornecer a fórmula sem limitação; e afirma benefícios de saúde vagos e amplos que, apesar de razoáveis, não foram examinados no estudo da FLP.

2a. Este é um estudo de coorte para avaliar se assistir a jogos de luta na

televisão prediz brigas futuras entre estudantes do ensino médio em Winston-Salem.

2b. Este é um estudo de caso-controle para saber se a duração da amamentação está associada com risco reduzido de câncer de ovário em mulheres chinesas que amamentaram pelo menos um bebê.

2c. Este é um estudo transversal sobre a relação entre a ingesta autorrelatada de gordura saturada e a concentração de esperma em homens dinamarqueses examinados para o serviço militar.

2d. Este é um ensaio clínico randomizado aberto para saber se um curso breve de vancomicina, lavagem intestinal e infusão duodenal de fezes de doador melhora a taxa de cura em 10 semanas em adultos com diarreia recorrente por *C. difficile* em comparação com um regime-padrão de vancomicina com e sem lavagem intestinal.

Cada uma dessas quatro frases é uma descrição concisa que resume todo o estudo, observando o delineamento e os principais elementos da questão de pesquisa (variáveis principais e população). Por exemplo, no exercício 2a, o delineamento é um estudo de coorte, o preditor é assistir a lutas na televisão, o desfecho são as brigas, e a população é de estudantes de ensino médio de Winston-Salem.

Capítulo 2 Elaborando a questão de pesquisa e desenvolvendo o plano de estudo

1. O processo de partir da questão de pesquisa para o planejamento do estudo costuma ser iterativo. Pode-se começar com uma resposta tipo: "um estudo transversal para determinar se a depressão está associada ao estado de saúde em adultos jovens". A possibilidade de que "depressão" esteja relacionada ao "estado de saúde" é Interessante e Relevante, mas a questão é vaga demais para permitir avaliar se ela atende os outros critérios FINER (Factível, Inovadora e Ética). Como a depressão e o estado de saúde serão medidos, e em que população? Também será difícil estabelecer causalidade por meio de um delineamento transversal – a depressão leva a uma piora no estado de saúde ou vice-versa?

Um delineamento mais específico que poderia atender os critérios FINER (dependendo de como ele for organizado) seria "Um estudo de

coorte para determinar se a depressão em estudantes universitários do primeiro ano, medida pelo questionário CES-D, prediz o número de consultas clínicas realizadas pelo estudante no próximo ano".

2. No caso da associação entre paracetamol e asma, a observação de que o uso de paracetamol e a prevalência de asma aumentaram no mundo todo (e a plausibilidade biológica relacionada com a depleção de glutationa reduzida pelo paracetamol) faz com que todos os estudos sejam Interessantes e Relevantes; à medida que mais estudos são realizados, eles se tornam menos Inovadores.

Estudo #1: Um estudo de caso-controle para comparar a frequência autorrelatada do uso de paracetamol em adultos com sintomas de asma vistos em clínicas de atenção primária no Sul de Londres (os casos), com a frequência relatada por adultos aleatoriamente selecionados sem tais sintomas nas mesmas clínicas de atenção primária (os controles). Estudos de caso-controle costumam ser uma boa forma de começar a investigação sobre possíveis associações (Capítulo 8). Este estudo foi especialmente Factível, pois foi parte de um estudo de caso-controle maior de base populacional que já investigava o papel de antioxidantes da dieta na asma. As razões de chances para asma aumentaram junto com a frequência do uso do paracetamol para até 2,38 (IC 95% 1,22 a 4,64) entre usuários diários (*P* para tendênciα = 0,0002). O estudo foi Ético, pois se tratou de estudo observacional que não colocou os sujeitos em risco (Shaheen et al. *Thorax* 2000;55:266-270).

Estudo #2: Um estudo transversal multinacional de sintomas alérgicos (asma, rinite alérgica e eczema) relatados pelos pais entre crianças de 6 a 7 anos, incluindo questões sobre o uso de paracetamol no ano anterior e o uso habitual para febre no primeiro ano de vida. Este estudo (que incluiu 205.487 crianças com idade entre 6 e 7 anos de 73 centros em 31 países) não seria factível se não fosse parte do estudo mais geral International Study of Asthma and Allergies in Childhood (ISAAC). Isso ilustra a importância de buscar dados existentes ou estudos existentes ao investigar uma nova questão de pesquisa (Capítulo 13). Os autores encontraram uma forte relação dose-resposta entre o uso atual de paracetamol e sibilância, com uma razão de chances de 1,46 (IC 95% 1,36-1,56) para sibilância e uma resposta "sim" para a questão: "Nos primeiros 12 meses de vida do seu filho, você costumava

administrar paracetamol para febre?" (Beasley et al. *Lancet* 2008;372:1039-1048).

Estudo #3: Um ensaio clínico randomizado duplo-cego sobre o efeito do paracetamol (12 mg/kg) versus ibuprofeno (5 ou 10 mg/kg) sobre hospitalizações e consultas ambulatoriais por asma ao longo de 4 semanas entre crianças febris com idade entre 6 meses e 12 anos que estavam sendo tratadas para asma no início do estudo. Um ensaio clínico randomizado costuma ser o delineamento clínico menos factível devido ao custo e logística envolvidos. Além disso, à medida que as evidências de um potencial efeito adverso do fármaco se acumulam, ensaios clínicos randomizados para confirmar isso se tornam menos éticos. Nesse caso, os investigadores realizaram uma análise retrospectiva dos dados entre as crianças com asma no Boston University Fever Study, um ensaio clínico randomizado duplo-cego que completou o arrolamento em 1993. Eles concluíram que as crianças randomizadas para o paracetamol tinham um risco 59% maior de hospitalização por asma (NS) e um risco 79% maior de consulta ambulatorial por asma (RR = 1,79, IC 95%: 1,05, 2,94; *P* = 0,01) (Lesko et al. *Pediatrics* 2002; 109:E20).

Capítulo 3 Escolhendo os sujeitos do estudo:especificação, amostragem e recrutamento

1a. Esta amostra de estudantes do terceiro ano pode não ser apropriada para a questão de pesquisa se os antecedentes do fumo ocorrerem mais cedo – pode ser melhor estudar alunos mais jovens. Além disso, a amostra do estudo (alunos de uma determinada escola de ensino médio) pode não ser representativa da população-alvo; os motivos que levam uma pessoa a fumar diferem em contextos culturais diversos. Seria melhor recrutar os sujeitos de várias escolas de ensino médio selecionadas aleatoriamente de toda a região. E, o que é o mais importante, o esquema de amostragem (chamar voluntários) atrai alunos que não representam o comportamento de fumo de toda a população acessível.

1b. A não representatividade da amostra poderia ter sido causada por

erro aleatório, embora isso seja pouco provável, a não ser que a amostra tenha sido muito pequena. Se a amostra fosse de 10, uma desproporção de 7:3 resultante do acaso ocorreria frequentemente; de fato, a probabilidade de se selecionar pelo menos 7 meninas em uma classe com 50% de meninas é de 17% (mais outros 17% de chance de selecionar pelo menos 7 meninos). Mas se o tamanho de amostra fosse de 100, a probabilidade de se selecionar pelo menos 70 meninas seria de menos de 0,01%. Isso ilustra o fato de que é possível estimar a magnitude do componente aleatório do erro amostral uma vez obtida a amostra e que é possível reduzir esse erro para qualquer valor desejado simplesmente aumentando o tamanho de amostra.

A amostra não representativa poderia também ter sido causada por erro sistemático. A grande proporção de estudantes do sexo feminino poderia ter sido causada pela diferença nas taxas de participação entre alunos e alunas. Estratégias para prevenir o viés de não resposta incluem as técnicas discutidas no Capítulo 3 para aumentar o recrutamento. A grande proporção de alunas também poderia representar um erro técnico ao se enumerar ou selecionar os nomes a serem amostrados. As estratégias para prevenção de erros incluem uso apropriado de pré-testes e procedimentos de controle de qualidade (Capítulo 17).

2a. Amostra aleatória (probabilística). A principal preocupação com a generalização será a não resposta – será importante fazer um questionário curto e fornecer algum incentivo para que seja preenchido. (O possível viés de não resposta é um problema em todos os esquemas de amostragem discutidos nesta questão.)

2b. Amostra aleatória estratificada (probabilística), em que mulheres entram em uma proporção três vezes maior, provavelmente porque se antecipou que haveria poucas mulheres no *show*.

2c. Amostra sistemática (não probabilística). Apesar de parecer conveniente, este esquema de amostragem sistemática levaria a uma sub-representação de ambos os membros dos casais. Além disso, pelo menos teoricamente, o vendedor da bilheteria poderia manipular as pessoas que receberiam os ingressos com final 1.

2d. Amostra por conglomerados (probabilística). Isso pode ser conveniente, mas esse tipo de amostragem deve ser levado em conta

nas análises, pois as pessoas que sentam na mesma fila podem ser mais semelhantes entre si do que o público aleatoriamente selecionado. Isso pode ser um problema importante se a música for mais alta em algumas filas em relação a outras.

2e. Amostra consecutiva (não probabilística). As amostras consecutivas costumam ser uma boa escolha, mas as pessoas que chegam cedo aos *shows* podem ser diferentes daquelas que chegam mais tarde, de modo que seria preferível realizar várias amostras consecutivas selecionadas em diferentes momentos.

2f. Amostra de conveniência (não probabilística). Este esquema perderá sujeitos que compraram ingressos pelo correio. Além disso, as pessoas que vão aos *shows* em grupos podem ser mais ou menos representadas.

2g. Amostra de conveniência (não probabilística). Este esquema de amostragem não apenas é enviesado pelos caprichos do investigador, como pode resultar em não resposta quando os espectadores não ouvem o convite.

3a. A população-alvo (para a qual os autores quiseram generalizar) foi a população norte-americana de crianças com menos de 5 anos no período estudado. Sabemos disso porque os autores usaram dados nacionais para estimar a carga da doença pelo metapneumovírus humano (MPVH). É claro que seria de grande interesse generalizar também para o futuro, e muitos leitores fariam isso sem pensar. Porém, é importante considerar, sobretudo com doenças infecciosas que podem variar a cada ano, que a generalização além do período do estudo é uma inferência adicional potencialmente frágil.

3b. A população acessível (a população de onde foram retirados os sujeitos) era de crianças < 5 anos que viviam ao redor dos três locais do estudo (Cincinnati, Nashville e Rochester, NY) e recebiam cuidados nestes locais. Presume-se que essas cidades tenham sido selecionadas em função de sua proximidade com os investigadores. Não está claro o quanto que elas representam outras regiões dos Estados Unidos com relação à frequência da infecção pelo MPVH.

3c. O esquema de amostragem foi uma amostra de conveniência. A escolha dos dias da semana (que não foi especificada) poderia levar a algum viés se, por exemplo, os pais de crianças com sintomas respiratórios mais leves no final de semana aguardassem até segunda-

feira para levá-las ao médico e se os sintomas de MPVH fossem mais ou menos intensos do que de outras viroses. Nos dias em que os investigadores estavam arrolando sujeitos, eles podem ter tentado obter uma amostra consecutiva (também não especificada), que teria ajudado a controlar o viés de seleção. A razão para restrição a certos meses do ano não é fornecida, mas presume-se que os autores acreditavam que quase todos os casos de MPVH ocorreriam durantes estes meses.

3d. As observações foram agrupadas por áreas geográficas, de maneira que o agrupamento por cidades deveria ser levado em conta na análise estatística. Quanto mais diferentes são as estimativas entre as cidades, mais isso ampliaria os intervalos de confiança. De maneira intuitiva, isso faz sentido. Taxas muito diferentes entre as cidades levariam alguém a se perguntar quão diferente a estimativa teria sido se outras cidades tivessem sido incluídas, e esperaríamos ver esta incerteza refletida em um intervalo de confiança mais amplo.

Um nível mais sutil de agrupamento ocorre pelo ano. Novamente, se houver muita variação de ano para ano na incidência de MPVH, e se a ideia for a generalização para anos futuros (em vez de apenas estimar qual era a incidência nos anos estudados), o agrupamento por ano precisaria ser levado em conta do ponto de vista estatístico e a variação significativa de ano para ano na incidência também levaria a um intervalo de confiança mais amplo.

Capítulo 4 Planejando as aferições: precisão,acurácia e validade

1a. Dicotômica
1b. Contínua
1c. Dicotômica
1d. Discreta ordenada
1e. Nominal
1f. Discreta ordenada
1g. Ordinal
1h. Contínua
1i. Nominal

1j. Dicotômica
O poder estatístico aumenta com o uso de variáveis de desfecho com informação ordenada. Por exemplo, o maior nível de escolaridade tem mais poder estatístico do que ensino médio completo/menos que ensino médio. Da mesma forma, o uso de índice de massa corporal como desfecho contínuo ofereceria um poder estatístico muito maior (por conter muito mais informação) para a maioria das questões de pesquisa do que a simples presença ou ausência de obesidade. Uma escolha intermediária comumente usada é a variável ordinal normal/sobrepeso/obeso.

2a. Este é um problema de acurácia. Poderia ser explicado pelo fato de o observador não ter visualizado o peso corretamente (um segundo observador poderia verificar o resultado), mas é mais provável que a balança necessite ser calibrada.

2b. Este é um problema de precisão. A variabilidade excessiva poderia ser um erro do observador, mas é mais provável que a balança necessite de conserto.

2c. Essa situação leva à perda de acurácia e precisão. Perde-se acurácia porque, quando o observador segura o bebê, ele altera o peso observado; dependendo da técnica usada, isso pode aumentar o peso consistentemente ou diminuí-lo consistentemente. Esse problema com os sujeitos pode ser resolvido solicitando-se que a mãe passe algum tempo acalmando o bebê; uma alternativa seria medir o peso do pai ou da mãe com e sem o bebê ao colo e então subtrair a diferença.

2d. O problema é principalmente de precisão, pois o indicador da balança irá oscilar em torno do peso real (se a balança estiver bem calibrada). O problema se relaciona aos sujeitos, e a solução é a mesma que no exercício 2c.

2e. Este é um problema principalmente de precisão, pois os pesos dos bebês variam se comeram ou “molharam a fralda” antes do exame. O problema é de variabilidade do sujeito e pode ser controlado instruindo-se as mães a não alimentarem os bebês nas 3 horas anteriores ao exame e pesando-se todos os bebês sem roupa.

3a. Validade preditiva: Os escores de *burnout* previram um desfecho que poderíamos esperar que estivesse associado com o *burnout*.

3b. Validade aparente: Perguntar às pessoas com que frequência se

sentem exaustas parece uma abordagem razoável para avaliar o *burnout*.

3c. Validade de construto: Esta medida de *burnout* é responsiva às circunstâncias que esperaríamos que afetassem o *burnout*.

3d. Validade de critério: Esses dois itens estão em pleno acordo com uma medida-padrão bem aceita.

Capítulo 5 Preparando-se para estimar o tamanho de amostra: hipóteses e princípios básicos

1. Tamanho de amostra = número projetado de sujeitos em um estudo que se espera que serão necessários para o investigador poder detectar uma determinada magnitude de efeito (para níveis especificados de α e β).

Hipótese nula = formulação da hipótese de pesquisa indicando que não há diferença entre os grupos em comparação.

Hipótese alternativa = formulação da hipótese de pesquisa indicando que há diferença entre os grupos em comparação.

Poder estatístico = probabilidade de se detectar uma diferença estatisticamente significativa entre os grupos em comparação (a partir de um determinado tamanho de amostra e com um determinado nível de significância estatística) se a diferença real na população for igual à magnitude de efeito.

Nível de significância estatística = probabilidade pré-definida de se rejeitar a hipótese nula, quando ela é verdadeira.

Magnitude de efeito = diferença mínima que o investigador deseja detectar entre os dois grupos em comparação.

Variabilidade = a quantidade de variação em uma medida, em geral calculada como desvio-padrão ou erro-padrão da média.

2a. Nenhum. Este é um resultado estatisticamente significativo, e não há nada que sugira que represente um erro Tipo I.

2b. O tamanho de amostra foi pequeno e poucos sujeitos teriam desenvolvido câncer de pulmão durante o estudo. Esses resultados negativos quase certamente se devem a um erro Tipo II, sobretudo devido às inúmeras evidências de outros estudos que mostram que o

fumo causa câncer de pulmão.

2c. Não há evidências epidemiológicas ou fisiopatológicas prévias de que o álcool reduz o risco de desenvolver diabetes; este resultado provavelmente se deve a um erro Tipo I. O investigador poderia ter sido mais informativo: $P < 0,05$ poderia ser $P = 0,04$ ou $P = 0,001$; este último reduziria (mas não descartaria) a probabilidade de erro Tipo I.

Capítulo 6 Estimando o tamanho de amostra e o poder estatístico: aplicações e exemplos

1. H_0: não há diferença no índice de massa corporal entre casos e controles de câncer gástrico.

H_A (bilateral): há diferença no índice de massa corporal entre casos e controles de câncer gástrico. O índice de massa corporal é uma variável contínua, e a condição caso-controle é uma variável dicotômica; logo, deve-se usar um teste *t*.

Magnitude de efeito (E) = 1 kg/m^2

Desvio-padrão (DP) = 2,5 kg/m^2

E/DP = 0,4

De acordo o Apêndice 6A,

Se $\alpha = 0,05$, $\beta = 0,20$, são necessários 100 sujeitos por grupo.

Se $\alpha = 0,05$, $\beta = 0,10$, são necessários 133 sujeitos por grupo.

Se $\alpha = 0,01$, $\beta = 0,20$, são necessários 148 sujeitos por grupo.

Opcional: Se o investigador teve acesso a apenas 60 casos, somente uma das estratégias a seguir irá ajudar:

a. Usar uma variável contínua – o índice de massa corporal já está sendo medido como variável contínua.

b. Usar uma variável mais precisa – o peso e a altura são variáveis precisas, e o desvio-padrão do índice de massa corporal é composto quase inteiramente de variação interindivíduo, que não pode ser reduzida. A padronização cuidadosa das medidas de peso e altura para reduzir o erro de mensuração ainda seria uma boa ideia, mas

não é a melhor escolha.

c. Usar medidas pareadas – não se aplica; a “mudança” no índice de massa corporal não é relevante para essa situação.

d. Usar um desfecho mais comum – não se aplica.

e. Usar tamanhos de amostra desiguais – o *n* dos controles pode ser aumentado, pois é fácil encontrar sujeitos sem câncer gástrico. Por exemplo, se o número de controles puder ser aumentado 4 vezes para um valor de 240, é possível usar a fórmula para aproximação encontrada no Capítulo 6, seção Estratégias para minimizar o tamanho de amostra e maximizar o poder, tópico Usar tamanhos desiguais para os grupos:

$$n' = ([c + 1] \div 2c) \times n$$

onde n' representa o “novo” número de casos, c representa a razão controle:caso (nesse exemplo, 4) e n representa o número “antigo” de casos (pressupondo um controle por caso). Neste exemplo,

$$n' = ([4 + 1] \div 8) \times 100 = (5 / 8) \times 100 = 63,$$

que é quase o número de casos disponíveis. Portanto, um estudo com 60 casos e 240 controles terá um poder estatístico semelhante ao de um estudo com 100 casos e 100 controles.

2. H_0: não há diferença na média de força muscular entre o grupo que recebeu DHEA e o grupo que recebeu placebo.

H_A: há diferença na média de força muscular entre o grupo que recebeu DHEA e o grupo que recebeu placebo.

$\alpha = 0,05$ (bilateral); $\beta = 0,10$

Teste = teste *t*

Magnitude de efeito = 10% × 20 kg = 2 kg

Desvio-padrão = 8 kg

A magnitude padronizada de efeito (E/DP) é 0,25 (2 kg/8 kg). No Apêndice 6A, partindo de 0,25 na coluna da esquerda e de α (bilateral) = 0,05 e $\beta = 0,10$, conclui-se que seriam necessários aproximadamente 338 sujeitos por grupo. Se for usado um $\beta = 0,20$, o tamanho de amostra passa a ser de 253 sujeitos por grupo.

3. H_0: não há diferença na média da mudança na força muscular entre o grupo que recebeu DHEA e o grupo que recebeu placebo.

H_A: há diferença na média da mudança na força muscular entre o grupo que recebeu DHEA e o grupo que recebeu placebo.

α = 0,05 (bilateral); β = 0,10

Teste = teste *t*

Magnitude de efeito = 10% × 20 kg = 2 kg

Desvio-padrão = 2 kg

A magnitude padronizada de efeito (E/DP) é 1,0 (2 kg/2 kg). No Apêndice 6A, partindo de 1,0 na coluna da esquerda e de α (bilateral) = 0,05 e β = 0,10, conclui-se que seriam necessários aproximadamente 23 sujeitos por grupo.

4. H_0: não há diferença na frequência de canhotos entre estudantes disléxicos e não disléxicos.

H_A: há diferença na frequência de canhotos entre estudantes disléxicos e não disléxicos.

α = 0,05 (bilateral); β = 0,20

Teste = teste do qui-quadrado (ambas as variáveis são dicotômicas)

Magnitude de efeito = razão de chances de 2,0

Sabendo que a proporção de estudantes não disléxicos canhotos (P_2) é de aproximadamente 0,1, o investigador quer detectar uma proporção de estudantes disléxicos canhotos (P_1) que produza uma razão de chances de 2,0. A estimativa do tamanho de amostra é feita com um teste do qui-quadrado, e usa-se o Apêndice 6B. No entanto, o apêndice foi criado pressupondo que seriam usadas as duas proporções, não a razão de chances. Apenas uma das proporções é conhecida (P_2 = 0,1).

Para calcular o valor de P_1 que fornece uma razão de chances de 2, pode-se usar a fórmula do Capítulo 6, seção Técnicas de tamanho de amostra para estudos analíticos e experimentais, tópico Teste do qui-

quadrado.

$$P_1 = \text{RC} \times P_2 \div ([1 - P_2] + [\text{RC} \times P_2]).$$

Neste exemplo:

$$P_1 = (2 \times 0{,}1) \div ([1 - 0{,}1] + [2 \times 0{,}1]) = 0{,}18$$

Portanto, P_1 é 0,18 e P_2 é 0,1. $P_1 - P_2$ é 0,08.

A Tabela 6B.2 no Apêndice 6B revela um tamanho de amostra de 318 indivíduos por grupo.

Opcional: Tente usar a fórmula do Capítulo 6, Seção Fórmula geral para outros valores; tenha persistência e use 6 casas decimais. Então obtenha uma resposta instantânea usando a calculadora na página deste livro na internet (em inglês): *www.epibiostat.ucsf.edu/dcr*.

5. O desvio-padrão dos escores de QI é cerca de um quarto da faixa "usual" (que é 170 – 130 = 40 pontos), ou seja, 10 pontos.

Amplitude total do intervalo de confiançα = 6 (3 acima e 3 abaixo). Nível de confiançα = 99%

Amplitude padronizada do intervalo de confiançα = amplitude total (A)/desvio-padrão (DP)

$$\text{A/DP} = 0{,}6$$

Na Tabela 6D, partindo de A/DP = 0,60 e de um nível de confiança de 99%, seriam necessárias médias de aproximadamente 74 escores de QI para se obter um escore médio com o intervalo de confiança especificado.

Capítulo 7 Delineando estudos transversais e de coorte

1a. Medem-se os níveis séricos de vitamina B_{12} em uma coorte de pessoas com mais de 70 anos de idade e sem história de fraturas de bacia, as quais são acompanhadas por um período de tempo (p. ex., 5 anos) para a ocorrência de fraturas, e depois se analisa a associação entre os níveis de vitamina B_{12} e a incidência de fraturas de bacia. (Poderia ser realizado um estudo menor, embora menos generalizável, analisando apenas mulheres, as quais têm maior taxa de fratura de bacia; um estudo ainda menor analisaria apenas mulheres brancas, as

quais têm as maiores taxas de fraturas.)

1b. Uma vantagem do delineamento de coorte prospectivo para estudar a associação entre vitamina B_{12} e fraturas de bacia:

- A sequência temporal (i.e., a fratura de bacia segue a deficiência de vitamina B_{12}) ajuda a estabelecer uma relação causa-efeito. Pessoas que fraturam a bacia podem desenvolver deficiência de vitamina B_{12} após a fratura porque diminuem a ingesta alimentar de vitamina B_{12} talvez em função da ida para uma clínica geriátrica.

Uma desvantagem do delineamento de coorte prospectivo:

- Um estudo de coorte prospectivo requer muitos sujeitos seguidos por muitos anos. O estudo, portanto, apresenta alto custo e demora muito para ser completado.

1c. Seria possível fazer um estudo de coorte retrospectivo se fosse encontrada uma coorte com soro armazenado e nessa coorte tivesse sido feito um seguimento relativamente completo para determinar quem desenvolveu fratura de bacia. A principal vantagem desse delineamento é que ele consome menos tempo e apresenta menor custo. A principal desvantagem é que o armazenamento do soro pode alterar os níveis de vitamina B_{12} e que as medidas de potenciais confundidores (p. ex., atividade física, fumo, etc.) podem não estar disponíveis.

2a. Embora o estudo PRIDE seja um ensaio clínico randomizado, o relato do exame basal é de um estudo transversal (observacional). Os estudos transversais costumam ser a primeira etapa em estudos de coorte ou ensaios clínicos randomizados.

2b. Apesar de ser possível que a depressão aumente a incontinência urinária, parece pelo menos igualmente plausível que a incontinência urinária aumente o risco de depressão. Conforme será discutido no Capítulo 9, também é possível que a associação seja causada por viés (por exemplo, se mulheres deprimidas tivessem maior probabilidade de *relatar* episódios de incontinência mesmo se não apresentassem maior frequência desses episódios), ou por confundimento, se um terceiro fator (p. ex., grau de obesidade) causasse a depressão e a

incontinência.

Um estudo longitudinal (coorte) ajudaria a esclarecer a sequência temporal da associação. Por exemplo, mulheres deprimidas e não deprimidas com pouca ou nenhuma incontinência no início do estudo poderiam ser acompanhadas para ver se aquelas deprimidas desenvolviam mais ou piorariam a incontinência com o tempo. Da mesma forma, mulheres continentes e incontinentes sem história de depressão poderiam ser acompanhadas para determinar se aquelas mais incontinentes tinham maior probabilidade de ficar deprimidas. Por fim, e de maneira mais convincente, os investigadores poderiam estudar as *alterações* na depressão ou incontinência, de ocorrência natural ou (idealmente) como resultado de uma intervenção, observando se as alterações em uma delas precediam as alterações na outra. Por exemplo, os sintomas depressivos melhoram quando a incontinência é tratada com sucesso? A incontinência (relatada) melhora com o tratamento da depressão?

Capítulo 8 Delineando estudos de caso-controle

1a. O grupo de casos pode ser composto de todas as mulheres entre 30 e 75 anos de idade com câncer de ovário relatado a um registro local de tumores, as quais poderiam ser contatadas por telefone e que concordassem em participar.

1b. Os controles podem ser uma amostra aleatória de todas as mulheres entre 30 e 75 anos de idade das mesmas cidades no registro de tumores. A amostra aleatória poderia ser obtida por meio de discagem aleatória (daí a necessidade de restringir os casos às pessoas que têm telefone).

1c. Como o câncer de ovário exige tratamento intensivo e pode ser fatal, é possível que alguns casos não estejam dispostos a participar do estudo ou já estejam mortos. Se a história familiar de câncer de ovário estiver relacionada somente a formas agressivas desse câncer, você poderá acabar subestimando o risco relativo, pois os casos com história familiar positiva têm menor probabilidade de sobreviverem por tempo suficiente para serem incluídos na amostra de casos. Se o

câncer de ovário familiar for mais benigno do que outros cânceres de ovário, poderia ocorrer o contrário.

Da mesma forma, é possível que mulheres saudáveis que têm um familiar com câncer de ovário estejam mais interessadas no estudo, e assim terão maior probabilidade de entrarem no estudo como controle. Nesse caso, a prevalência de história familiar de câncer de ovário no grupo-controle será artificialmente elevada, e a estimativa do risco para câncer de ovário devido à história familiar será falsamente baixa. É possível minimizar esse problema, se não for informado aos potenciais controles exatamente qual é a questão de pesquisa ou exatamente que câncer está sendo estudado, desde que isso seja feito de forma aceitável para o Comitê de Ética em Pesquisa (CEP).

1d. A história familiar de câncer de ovário costuma ser medida perguntando-se aos sujeitos quantos familiares do sexo feminino eles têm e quantos tiveram essa doença. O problema principal dessa abordagem é o viés recordatório. Mulheres com câncer de ovário que podem estar preocupadas com a possibilidade de predisposição genética para a sua doença têm maior tendência a se lembrar de algum familiar que tinha câncer de ovário do que mulheres sem razões para pensar nessa possibilidade. Nesse caso, a estimativa da associação entre história familiar e câncer de ovário pode ser falsamente elevada.

Além disso, muitas mulheres confundem as neoplasias ginecológicas (da cérvice uterina, do corpo uterino e do ovário) e podem confundir tumores ginecológicos benignos que requerem cirurgia com tumores malignos. Isso pode gerar classificações errôneas (mulheres sem história familiar de câncer do ovário relatam ter o fator de risco e são classificadas inapropriadamente). Quando esse problema ocorre em igual grau em casos e controles, a estimativa da associação entre história familiar e câncer do ovário é falsamente baixa. Quando esse problema é mais comum nos casos (que podem ter maior tendência a interpretar erroneamente o tipo de câncer ou a razão para cirurgia nos familiares), então a estimativa da associação entre história familiar e câncer de ovário é falsamente elevada. O problema da classificação errônea pode ser reduzido verificando-se o diagnóstico em registros patológicos de membros da família que foram apontados como tendo tido câncer de ovário para confirmar o diagnóstico.

Por fim, seria desejável levar em conta a *oportunidade* para que casos e controles tenham uma história familiar positiva: mulheres com muitas irmãs mais velhas têm mais oportunidade de apresentar história familiar positiva do que aquelas apenas com irmãos ou irmãs mais jovens. Conforme discutido no Capítulo 9, o pareamento e a estratificação são duas maneiras de lidar com essa possibilidade.

1e. A abordagem mais simples seria dicotomizar a história familiar de câncer de ovário (p. ex., parentes de primeiro grau ou não) e usar a razão de chances como medida de associação. A razão de chances aproxima o risco relativo porque o desfecho (câncer de ovário) é raro. Um teste simples do qui-quadrado seria o teste apropriado de significância estatística. Outra opção, caso a história familiar seja quantificada (p. ex., proporção de familiares do sexo feminino de primeiro e segundo grau afetados), seria procurar uma dose-resposta, computando a razão de chances para cada nível da exposição.

1f. O delineamento de caso-controle é uma forma razoável de se responder a essa questão de pesquisa, apesar dos vieses de amostragem, de recordação e de classificação. A principal alternativa seria um grande estudo de coorte, mas, como o câncer do ovário é muito raro, o estudo de coorte provavelmente não seria factível. Um estudo de coorte retrospectivo no qual os dados de história familiar já tivessem sido sistematicamente coletados seria ideal, se essa coorte pudesse ser encontrada.

2a. Os casos poderiam ser os motoristas mais jovens (talvez 16 a 20 anos) envolvidos em acidentes, e os controles poderiam ser amigos ou conhecidos. Seria importante excluir amigos com os quais eles jogam *videogame* para evitar pareamento excessivo. É provável que a discagem aleatória tenha menos sucesso como estratégia para identificar os controles devido à alta prevalência de telefones celulares (os quais, diferentemente dos telefones fixos, não são geograficamente localizados) nessa faixa etária. Os casos e controles também poderiam ser identificados se o investigador tivesse acesso aos registros de uma companhia de seguros automotivos. Poder-se-ia argumentar que casos e controles deveriam ser pareados quanto ao sexo, pois tanto o hábito de jogar *videogame* como os acidentes automobilísticos são mais comuns em homens jovens. A exposição poderia ser medida

utilizando-se um questionário ou entrevista sobre o uso de *videogames*. Seria importante perguntar sobre *videogames* que não envolvem direção, bem como aqueles que envolvem, pois a inferência causal seria aumentada se a associação fosse específica, ou seja, se houvesse um efeito para o uso de jogos de direção/corrida, mas não para jogos de tiros ou de outros tipos.

2b. Quando se acredita que exposições intermitentes têm um efeito a curto prazo, como o uso de um *videogame* logo antes de dirigir, um estudo cruzado de caso é uma boa opção. Conforme descrito no exercício 2a, os casos poderiam ser os motoristas mais jovens envolvidos em acidentes. Em um estudo cruzado de caso, não há controles, apenas os períodos de tempo controlados. Assim, os motoristas dos casos seriam questionados sobre o uso de *videogames* de corridas logo antes de seu envolvimento em acidentes e também sobre períodos de tempo controlados em que não se envolveram em acidentes. O período de tempo logo antes do acidente é comparado em uma análise pareada com outros períodos de tempo para ver se o uso de *videogames* de corrida era mais comum no período antes do acidente em comparação com outros períodos de tempo.

Capítulo 9 Fortalecendo a inferência causal em estudos observacionais

1. Há cinco explicações possíveis para a associação entre a ingesta de frutas, verduras e legumes na dieta e doença coronariana:
   a. Acaso – o achado de que indivíduos com doença coronariana comem menos frutas, verduras e legumes deve-se a erro aleatório. Como discutido no Capítulo 5, o valor *P* permite quantificar a magnitude da diferença observada em comparação com a que poderia ser esperada pelo acaso; o intervalo de confiança de 95% mostra a variação de valores consistente com os resultados do estudo. Mantendo-se todo o resto igual, quanto menor o valor *P* e quanto mais distante o valor nulo está da extremidade mais próxima do intervalo de confiança, menos plausível é a possibilidade do acaso como explicação.
   b. Viés – houve erro sistemático (diferença entre a questão de

pesquisa e a forma como o plano de estudo foi conduzido) na amostra, na variável preditora ou na variável de desfecho. Por exemplo, a amostra pode apresentar viés se os controles forem todos pacientes de convênio que compareceram a uma revisão anual de saúde, pois esses pacientes podem ser mais conscientes sobre sua saúde (e assim comer mais frutas, verduras e legumes) do que toda a população em risco de doença coronariana. As medições de aspectos da dieta podem apresentar viés se os indivíduos que sofreram um infarto tenderem a recordar melhor os padrões inadequados de dieta do que os controles (viés recordatório) ou se entrevistadores não cegos fizerem as perguntas ou registrarem as respostas de forma diferente para casos e controles.

c. Efeito-causa – é possível que o infarto tenha causado alterações nas preferências dietéticas do indivíduo, de forma que comessem menos frutas, verduras e legumes do que antes do infarto. A possibilidade de uma relação efeito-causa pode ser abordada delineando-se variáveis para examinar a sequência histórica – por exemplo, perguntando-se aos casos e controles sobre sua dieta anterior, em vez da atual.

d. Confundimento – pode haver outras diferenças entre os indivíduos que comem mais frutas, verduras e legumes e os indivíduos que comem menos desses alimentos, e tais diferenças podem ser a real causa da taxa mais baixa de doença coronariana. Por exemplo, indivíduos que comem mais frutas, verduras e legumes podem praticar mais atividade física.

As possíveis estratégias para controlar os fatores de confusão causados pela atividade física são apresentadas na tabela a seguir:

Método	Plano	Vantagens	Desvantagens
Fase de delineamento			
Especificação	Arrolar somente indivíduos que relatam não praticarem atividade física regular.	É uma abordagem simples.	Limita o número de sujeitos elegíveis, dificultando o recrutamento. O estudo pode não ser generalizável para indivíduos que praticam atividade física.
Pareamento	Parear cada caso com um controle com	Elimina o efeito da atividade física como	Exige um maior empenho para identificar controles que possam ser pareados com

	níveis semelhantes de atividade física.	preditor da doença coronariana, podendo aumentar levemente a precisão (poder estatístico) para observar a dieta como variável preditora.	cada caso. Podem-se desperdiçar casos se não forem encontrados controles com níveis semelhantes de atividade física. Elimina a possibilidade de estudar o efeito da atividade física na doença coronariana.
Fase de análise			
Estratificação	Ao fazer a análise, agrupar os sujeitos em três ou quatro estratos de acordo com os níveis de atividade física.	Fácil de executar, fácil de entender e reversível.	Só faz sentido avaliar um número pequeno de estratos e de variáveis confundidoras. Perde-se parte da informação contida na variável contínua condicionamento físico quando essa variável é convertida em categórica, e isso pode resultar em um controle incompleto do confundimento.
Ajuste estatístico (modelagem)	Usar um modelo de regressão logística para controlar para o condicionamento físico e para outros potenciais confundidores.	Pode controlar reversivelmente para toda a informação contida na variável condicionamento físico como variável preditora contínua, controlando ao mesmo tempo para outros potenciais confundidores, como idade, raça e fumo.	O modelo estatístico pode não se adequar aos dados, resultando em um controle incompleto do confundimento e em resultados potencialmente enganosos. Por exemplo, o efeito da dieta ou do condicionamento físico pode não ser o mesmo em fumantes e não fumantes. Os potenciais confundidores importantes devem ser medidos antecipadamente. Às vezes é difícil entender e descrever os resultados do modelo, sobretudo quando as variáveis não são dicotômicas.

Além dessas quatro estratégias para controlar o confundimento em estudos observacionais, há uma solução mais definitiva: delinear um ensaio clínico randomizado.

e. Causa-efeito – a quinta explicação possível é que comer frutas, verduras e legumes realmente reduz a taxa de eventos coronarianos. Essa explicação torna-se provável, em parte, pela exclusão de cada uma das outras quatro explicações possíveis e, em parte, pela busca de outras evidências para apoiar a hipótese causal. Um exemplo dessa última estratégia é considerar as evidências biológicas de que algum componente de frutas, verduras e legumes (p. ex., um antioxidante) seja antiaterogênico, e estudos ecológicos concluindo

que a doença coronariana é muito menos comum em populações que comem mais frutas, verduras e legumes.

2. Este é um exemplo de condicionamento em um efeito comum: O estudo incluiu apenas lactentes com febre, a qual pode ser causada por infecções urinárias e de ouvido. Visto que meninos não circuncisados tinham maior probabilidade de ter uma infecção do trato urinário, eles apresentavam maior probabilidade de ter outra causa para sua febre que não infecção de ouvido (i.e., eles foram excessivamente representados entre os meninos que não tinham infecção de ouvido).
3. A associação entre o uso materno de paracetamol e asma na prole poderia ser examinada em um estudo de coorte no qual as mães seriam questionadas sobre o uso de paracetamol durante a gestação e os filhos seriam acompanhados quanto ao desenvolvimento de asma. Os investigadores procurariam evidências de que o genótipo materno modifique o efeito da exposição materna ao paracetamol sobre asma em crianças (interação), com uma associação mais forte entre exposição e desfecho entre aquelas mais geneticamente suscetíveis. De fato, isso é o que foi relatado por Shaheen et al. (*J Allerg Clin Immunol* 2010;126(6):1141-1148 e7.) no Avon Longitudinal Study of Parents and Children (ALSPAC).

Capítulo 10 Delineando um ensaio clínico randomizado cego

1a. A principal vantagem de se usar biomarcadores (uma variável contínua) como desfecho primário em um ensaio clínico é um menor tamanho de amostra e uma menor duração do estudo para determinar se o tratamento reduz o nível do marcador. A principal desvantagem é a incerteza sobre se a mudança nos níveis do marcador induzida pelo tratamento significa que o tratamento irá reduzir a incidência do desfecho muito mais importante clinicamente (desenvolvimento de demência).

1b. O diagnóstico clínico de demência é um desfecho com maior significado que poderia melhorar a prática clínica de prevenção da demência. A desvantagem é que um ensaio clínico desse tipo seria grande, longo e caro.

2a. Deve-se perguntar aos participantes, em cada visita de seguimento, se eles apresentam diarreia, náusea ou vômitos. Isso pode ser feito solicitando-se que os sujeitos assinalem seus sintomas em quadros de verificação que são de formato fácil de codificar e analisar. Para descobrir outros efeitos adversos não antecipados, os participantes deveriam também descrever outros sintomas, condições ou cuidados médicos (como hospitalização ou novos medicamentos que necessitam de receita médica) que ocorreram após a última visita. Isso deve ser perguntado de forma aberta, mas as respostas podem ser codificadas para a entrada dos dados.

2b. A coleta de dados na linha de base deve incluir: (1) informações sobre como contatar o participante, a família ou o médico para permitir um seguimento mais completo; (2) características da população arrolada (como idade, etnia/raça e gênero) para permitir a descrição da coorte do estudo; (3) fatores de risco para o desfecho (como hipertensão ou história familiar de demência) que poderiam identificar participantes com a maior taxa do desfecho e poderiam ser usados para demonstrar que os grupos do estudo eram comparáveis na linha de base, bem como definir os subgrupos para análises secundárias; e (4) medida dos desfechos (intensidade do déficit cognitivo). Amostras biológicas devem ser armazenadas para permitir medição futura de fatores, como genótipos de enzimas que metabolizam o fármaco, que poderiam influenciar a efetividade do tratamento.

2c. A randomização em blocos e estratificada poderia garantir um número muito semelhante de participantes com o genótipo Apo«4 nos grupos tratado e controle. Isso pode ser especialmente importante se o efeito do tratamento for influenciado pela presença do genótipo. Por outro lado, esse processo torna o ensaio clínico mais complicado (avaliar o genótipo Apo«4 antes do arrolamento irá postergar a randomização e levantar questionamentos relacionados a como aconselhar os participantes sobre os resultados). O risco de um desbalanço substancial em um ensaio clínico relativamente grande (> 200 pacientes por braço do estudo) é muito baixo, de forma que a randomização simples seria uma boa escolha.

Capítulo 11 Delineamentos alternativos para o ensaio clínico e tópicos relacionados à implementação

1. O principal objetivo de um ensaio clínico de fase I é determinar se o tratamento é suficientemente seguro e bem tolerado para permitir que ensaios clínicos posteriores encontrem a melhor dose e testem sua efetividade. Um ensaio clínico de fase I arrolaria homens com alopecia masculina e usaria uma ou mais potenciais doses do tratamento (escalonando a dose apenas se a dose anterior não causasse efeitos colaterais), com o desfecho principal sendo a ocorrência de eventos adversos, como *rash* cutâneo. Não haveria grupo-controle.

2a. O valor da comparação entre finasterida e HairStat depende, em grande parte, da força dos dados para sustentar o uso da finasterida como padrão de cuidados para o tratamento da alopecia masculina. Se esses dados não forem muito fortes, ou se a finasterida não for comumente usada na prática clínica, seria melhor comparar o HairStat com placebo. Um ensaio clínico controlado por placebo forneceria evidências claras de que o HairStat é melhor que placebo. Seria razoável comparar o HairStat com finasterida se a finasterida for considerada o padrão de cuidados para a alopecia masculina e se houver ensaios clínicos randomizados de boa qualidade documentando a eficácia da finasterida. Nesse caso, os investigadores devem primeiro decidir se consideram o HairStat mais efetivo do que a finasterida. Se for assim, um estudo de comparação ativa seria a melhor escolha para comparar HairStat com finasterida. Se os investigadores considerarem que o HairStat é tão bom quanto a finasterida, mas é muito mais barato, eles devem considerar um estudo de não inferioridade. Nesse caso, eles devem ter cuidado para utilizar um delineamento que seja muito semelhante àquele usado para documentar a eficácia da finasterida (critérios de inclusão, dose, duração do tratamento, medidas de desfecho) e devem conduzir o estudo para garantir uma quantidade mínima de falta de adesão e perda de seguimento. Um grande problema dos estudos de não inferioridade é que o tamanho da amostra deve ser muito maior do que aquele necessário para um

estudo controlado por placebo.

2b. Um delineamento fatorial que inclui um placebo tem a vantagem de comparar cada tratamento com um placebo e (se planejado com poder estatístico suficiente) testar se a combinação de tratamentos é melhor do que qualquer um dos tratamentos isolados. A desvantagem óbvia é o maior tamanho, o custo e a complexidade do estudo.

3. A adesão às visitas, ao protocolo e à medicação do estudo pode ser melhorada adotando-se as seguintes medidas:

- empregar funcionários atenciosos, que estejam entusiasmados com o estudo;
- enviar lembretes (por e-mail, telefone ou correspondência) sobre as visitas seguintes e sobre a importância da adesão ao tratamento;
- oferecer reembolso para viagens, estacionamento e outros gastos relacionados ao estudo;
- considerar duas consultas de rastreamento antes da randomização para identificar os participantes com maior probabilidade de faltarem às visitas de seguimento;
- fazer um período de *run-in* durante o qual os participantes devem usar um gel capilar de placebo, excluindo-se aqueles sem adesão ao tratamento;
- outras possíveis estratégias listadas na Tabela 11.2.

4. A principal desvantagem da intenção-de-tratar é que ela inclui participantes que não aderem ao tratamento randomizado, reduzindo, assim, a magnitude aparente de qualquer efeito observado para todo o grupo randomizado. No entanto, as desvantagens de se usar a análise por protocolo no lugar da análise por intenção-de-tratar são ainda maiores – como os participantes que não aderem à intervenção normalmente diferem de forma importante daqueles que aderem, deixa de haver uma verdadeira comparação randomizada, podendo-se concluir erroneamente que o HairStat é efetivo.

5. A conclusão de que o HairStat funciona melhor em homens jovens com base em uma análise de subgrupos pode ser errônea e enganadora,

pois o resultado pode se dever ao acaso. A probabilidade de se encontrar um efeito “significativo” em um subgrupo, quando não há efeito significativo no grupo maior, aumenta com o número de subgrupos testados; não fica claro quantos subgrupos foram testados para encontrar esse efeito “significativo”. O argumento de que o tratamento é efetivo em homens com menos de 40 anos implica que o tratamento não foi eficaz – ou até mesmo teve um efeito oposto – em homens mais velhos. Esse resultado também deve ser relatado e testado estatisticamente para uma modificação do efeito de crescimento capilar do HairStat causada pela idade. O argumento de que o HairStat é efetivo no subgrupo de homens mais jovens só pode ser feito se a análise de subgrupos for especificada com antecedência (idealmente com base em aspectos biológicos para suspeitar que o HairStat poderia funcionar melhor em homens mais jovens), se não houve um grande número de subgrupos testados e se o valor de *P* para modificação do efeito (interação) entre o efeito do tratamento e a idade foi estatisticamente significativo.

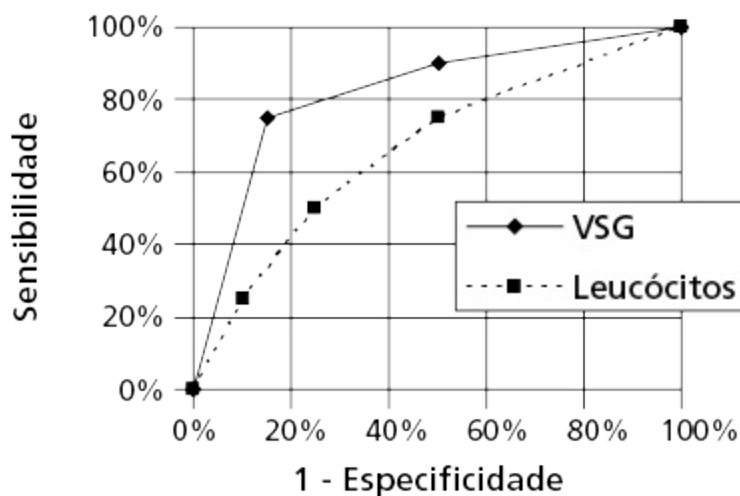
Capítulo 12 Delineando estudos sobre testes médicos

1a. Em geral a melhor forma de amostrar sujeitos para um teste diagnóstico é amostrar pacientes em risco de desenvolver uma doença, antes que se saiba quem tem e quem não tem a doença. Nesse caso, talvez a melhor escolha seja amostrar mulheres que vão a um serviço de pronto-atendimento ou de emergência com sinais ou sintomas compatíveis com doença inflamatória pélvica (DIP). Talvez a pior escolha seja comparar as velocidades de sedimentação globular (VSGs) de mulheres hospitalizadas com DIP com as de uma população-controle saudável, pois o espectro de doença e, principalmente, o espectro de não doença não são representativos dos grupos em que os testes seriam usados na prática clínica. (Mulheres hospitalizadas com DIP provavelmente têm doença mais grave do que a média, e mulheres voluntárias em bom estado de saúde têm uma probabilidade muito menor de ter VSGs elevadas do que mulheres com dor abdominal resultante de causas que não a DIP.)

1b. Se os indivíduos responsáveis pelo diagnóstico final usaram a VSG para decidir quem tinha e quem não tinha DIP, isso pode ter aumentado falsamente a sensibilidade e a especificidade. O viés resultante (chamado “viés de incorporação”) aumenta à medida que aumenta o papel da VSG na decisão diagnóstica.

1c. A melhor resposta é que não deveria ser usado um ponto de corte para definir um resultado como anormal. Dever-se-ia, em vez disso, criar um gráfico mostrando o contrabalanço (*trade-off*) entre sensibilidade e especificidade usando uma curva ROC e apresentar uma tabela das propriedades diagnósticas para vários intervalos de VSGs (p. ex., < 20, 20 a 49, ≥ 50 mm/h), em vez das sensibilidades e especificidades. Isso está ilustrado pela tabela a seguir, a qual pode ser criada a partir da informação na questão:

VSG	Com DIP	Sem DIP	Razão de verossimilhança
≥ 50	75%	15%	5,00
20–49	15%	35%	0,43
< 20	10%	50%	0,20
	100%	100%	

Pode-se também usar uma curva ROC para comparar a VSG com um ou mais testes como contagem de leucócitos. Isso é ilustrado pela curva ROC hipotética a seguir, que sugere que a VSG é superior à contagem de leucócitos para predizer DIP:

2a. Esse problema ilustra o erro comum que é excluir indivíduos do numerador, sem excluí-los do denominador. Embora seja verdade que houve apenas duas crianças com lesões intracranianas “inesperadas”, o denominador para o rendimento deveria ser o número de crianças nas quais as lesões intracranianas seriam consideradas inesperadas, isto é, aquelas com exame neurológico e estado mental normais – provavelmente um número muito menor do que 200. Por exemplo, suponha que apenas 50 daquelas encaminhadas para uma tomografia tenham achados normais no exame neurológico e no estado mental. Nesse caso, o rendimento seria de 2 para 50, ou 4% – quase quatro vezes maior.

2b. A não ser que o achado de lesão intracraniana resulte em mudanças no manejo dos pacientes e que haja uma forma de se estimar os efeitos dessas mudanças no desfecho, será muito difícil saber que valor de rendimento fará o teste valer a pena. Seria melhor usar “lesão intracraniana que requer intervenção” como desfecho nesse estudo, embora isso exija um certo consenso sobre que lesões requerem intervenção e uma certa estimativa da efetividade dessas intervenções na melhora do desfecho.

2c. A primeira vantagem é a capacidade de se examinar possíveis benefícios de resultados normais. Por exemplo, uma TC normal poderia mudar o plano de manejo de “colocar em observação” para “mandar para casa”. Em estudos sobre rendimento diagnóstico, em geral se pressupõe que resultados normais têm pouco valor. Uma segunda vantagem, como mencionado, é que TCs anormais podem não levar a nenhuma mudança no manejo dos pacientes (p. ex., se a neurocirurgia não era necessária e se os pacientes seriam internados mesmo assim). Estudar os efeitos de testes na tomada de decisões clínicas ajuda a determinar o quanto eles podem contribuir com novas informações além do que já se sabia ao se solicitar o teste.

3a. Se apenas forem incluídas crianças que haviam feito uma TC, o estudo será suscetível ao primeiro tipo de viés de verificação (Apêndice 12B), no qual a sensibilidade é falsamente elevada e a especificidade é falsamente diminuída, pois crianças sem alterações neurológicas focais (“falso-negativos” ou “verdadeiro-negativos”) estarão sub-representadas no estudo.

3b. Se crianças com traumatismos craniencefálicos que não fizeram uma TC forem incluídas, e for pressuposto que elas não apresentaram lesão intracraniana caso tenham se recuperado sem neurocirurgia, o estudo estará suscetível ao segundo tipo de viés de verificação, "o viés do duplo padrão-ouro" (Apêndice 12C), que tende a aumentar tanto a sensibilidade como a especificidade, se algumas lesões intracranianas melhorarem espontaneamente sem necessidade de neurocirurgia.

Capítulo 13 Usando bases de dados existentes

1. Algumas possibilidades:
   a. Analisar dados do National Health and Nutrition Examination Survey (NHANES). Esses estudos nacionais são conduzidos periodicamente, e os seus dados estão disponíveis para qualquer investigador a um custo simbólico. Eles contêm dados de amostras de base populacional, incluindo variáveis sobre história clínica autorrelatada de doença da vesícula biliar e resultados de ecografia abdominal.
   b. Analisar dados do Medicare sobre a frequência de cirurgia de vesícula biliar em pacientes com mais de 65 anos de idade nos Estados Unidos ou dados do National Hospital Discharge Summary sobre a frequência dessa cirurgia em indivíduos de todas as idades. Ambas as bases de dados contêm uma variável sobre raça. Os denominadores podem vir de dados censitários. Assim como o NHANES, essas amostras de base populacional são muito boas, mas têm o problema de responder a uma questão de pesquisa um pouco diferente (quais são as taxas de tratamento cirúrgico para doença da vesícula biliar?). A resposta a essa questão pode diferir da incidência real de doença da vesícula biliar devido a fatores como acesso a cuidados médicos.

2a. As principais vantagens são que usar os dados do CHS em uma análise de dados secundários foi uma estratégia rápida, fácil e barata – sobretudo se comparada com o tempo e os gastos que estariam envolvidos no planejamento e na condução de um grande estudo de coorte. Além disso, o *fellow* de pesquisa desenvolveu desde então uma colaboração com os investigadores do CHS e conseguiu acrescentar

medidas mais sofisticadas de função renal ao CHS como um estudo suplementar.

2b. Em alguns casos, o banco de dados secundário não fornece as medidas ideais do preditor, do desfecho ou de confundidores. É importante se assegurar de que o banco de dados irá fornecer respostas razoáveis à questão de pesquisa antes de investir o tempo e o esforço necessários para obter acesso aos dados. Outro problema é que pode ser difícil obter dados de outros estudos – o investigador em geral precisa escrever uma proposta, encontrar um colaborador que seja coinvestigador do estudo e obter aprovação do Conselho Diretivo e do patrocinador.

3. Já foram realizados diversos ensaios clínicos randomizados de grande porte acerca do efeito do estrogênio e de SERMs sobre vários desfechos, incluindo eventos cardiovasculares, câncer e eventos tromboembólicos. Esses ensaios clínicos incluem o Women's Health Initiative, o Breast Cancer Prevention Trial, o Multiple Outcomes of Raloxifene Evaluation e o Raloxifene Use for the Heart. O melhor lugar para o investigador iniciar é determinando se o estrogênio pode ser medido a partir de soro armazenado congelado e, se puder, verificar se algum desses ensaios clínicos de grande porte dispõem de soro armazenado que poderia ser usado para fazer essa medição. O melhor delineamento para essa questão seria um estudo de caso-controle aninhado ou de caso-coorte aninhado. O investigador provavelmente irá precisar redigir uma proposta para esse estudo suplementar, obter aprovação do Conselho Diretivo e do patrocinador e obter financiamento para fazer as medições adicionais (o que envolverá um valor relativamente baixo, pois a maior parte dos custos já foi coberta pelo ensaio clínico principal).

Capítulo 14 Abordando questões éticas

1a. Isso depende de os participantes do estudo original terem dado ou não consentimento para o uso de suas amostras para sequenciamento de DNA para utilização em outros estudos futuros, e de que tipos de estudos foram especificados. O consentimento original não cobriria a pesquisa proposta se as amostras de sangue fossem coletadas apenas

para repetir os testes especificados no protocolo, em caso de perda de amostras ou acidentes no laboratório (como colesterol e hemoglobina A1c). Da mesma forma, o consentimento original não cobriria a pesquisa proposta se os participantes tivessem dado consentimento para que as amostras de sangue fossem usadas para mensuração genética de DNA em outros estudos sobre doença coronariana, mas sem menção sobre o uso dessas amostras em estudos sobre diabetes.

1b. Sob a norma federal, um projeto pode ser realizado em amostras e dados existentes se o novo investigador não puder identificar os participantes, seja diretamente ou com o auxílio de outra pessoa. Assim, se o novo pesquisador receber amostras e dados rotulados como 0001, 0002, 0003, etc., e o código que une as amostras com a identidade dos participantes tiver sido destruído ou não for acessível pelo novo pesquisador, não é necessário obter consentimento adicional para o estudo secundário. A justificativa ética é que o caráter anônimo dos materiais protege os participantes de quebras na confidencialidade, que é o maior risco na pesquisa que envolve materiais e dados existentes. Pressupõe-se que ninguém iria se opor a que seus materiais e dados fossem usados se não houvesse risco de quebra de confidencialidade. Deve-se lembrar, no entanto, que alguns participantes poderiam ter objeções a alguém sequenciar seu DNA, mesmo se a confidencialidade fosse mantida, pois o DNA contém informações que poderiam, futuramente, levar a uma perda de confidencialidade.

1c. Quando os pesquisadores coletam novas amostras para um projeto de pesquisa, é prudente pedir permissão para coletar e armazenar sangue adicional para ser usado em pesquisas futuras. Armazenar amostras permite que futuramente sejam realizados novos estudos de forma mais eficiente do que montando uma nova coorte. Recomendam-se estágios progressivamente mais abrangentes de consentimento: solicita-se ao participante consentimento (1) para o estudo específico (p. ex., o estudo de coorte original), (2) para outros projetos de pesquisa sobre o mesmo assunto geral (como risco de doença coronariana), ou (3) para todas as pesquisas futuras aprovadas por um CEP e por um painel de revisão científica. Para abordar as questões levantadas em 1b, pode-se também solicitar ao participante

consentimento específico para pesquisas envolvendo sequenciamento do seu DNA. O participante pode concordar com uma, algumas ou todas as opções. Obviamente, é impossível descrever pesquisas futuras. Portanto, o consentimento para estudos futuros não é realmente informado no sentido de que o participante não saberá a natureza, os riscos e os benefícios dos estudos futuros. O participante está sendo solicitado a confiar que os CEPs e os painéis de revisão científica irão permitir estudos futuros apenas quando eles atenderem a critérios científicos e éticos rigorosos.

2a. Negar ao grupo-controle fármacos sabidamente eficazes iria sujeitá-los a dano, e, portanto, seria antiético. Mesmo se os participantes fornecessem consentimento informado para participar em um ensaio clínico controlado por placebo desse tipo, um CEP não aprovaria tal estudo, que viola as exigências regulatórias de que o balanço risco/benefício deve ser aceitável, e os riscos, minimizados.

2b. Se todos os participantes no estudo fossem tratados com a quimioterapia-padrão atual, os participantes também poderiam ser randomizados para o novo tratamento ou placebo. De modo alternativo, os investigadores devem tentar identificar um subgrupo de pacientes para os quais nenhum tratamento comprovadamente prolonga a sobrevida (o desfecho mais clinicamente relevante na maioria dos tratamentos contra o câncer). Por exemplo, os pacientes cuja doença progrediu apesar de vários tipos de quimioterapia-padrão e que não dispõem de opções comprovadamente eficazes poderiam ser convidados a participar em um ensaio clínico controlado por placebo sobre a intervenção experimental. Um braço-controle aceitável pode ser placebo ou o melhor tratamento atual. Essa abordagem pressupõe que se o novo fármaco é ativo em pacientes previamente não tratados, ele também será ativo após outros tratamentos terem falhado. É obviamente possível que um novo fármaco que não funciona em uma doença refratária seja eficaz como tratamento de primeira linha.

3a. Durante o consentimento informado, os investigadores devem discutir: (1) a natureza do estudo; (2) o número e a duração das visitas; (3) a descrição dos potenciais benefícios e riscos da participação no estudo (nesse caso, principalmente o estigma e a discriminação caso haja quebra de confidencialidade); (4) alternativas à participação no

estudo, incluindo medidas de prevenção contra o HIV disponíveis fora do estudo; (5) a natureza voluntária da participação e o direito de abandonar o estudo a qualquer momento; e (6) a proteção da confidencialidade consistente com as exigências de notificação das autoridades de saúde pública.

3b. Os investigadores devem apresentar os dados de maneira que os participantes possam compreender. Os participantes com poucos conhecimentos sobre saúde não conseguirão compreender um termo de consentimento informado detalhado. Aconselha-se obter apoio de grupos ligados à comunidade sobre a melhor maneira de apresentar as informações. Algumas sugestões incluem fitas de vídeo, DVDs e revistas em quadrinhos. Pré-testes extensos devem ser realizados. Além disso, os pesquisadores devem procurar identificar quais concepções equivocadas que surgem sobre o estudo são mais comuns e revisar o processo do consentimento para poder abordá-las.

3c. Mesmo que o estudo seja observacional, os pesquisadores têm uma obrigação ética de fornecer informações aos participantes sobre como reduzir seu risco de contrair HIV. Há motivos éticos e científicos para isso. Os pesquisadores têm a obrigação ética de prevenir danos aos participantes de seu estudo. Eles não podem deixar de oferecer medidas factíveis de saúde pública sabidamente eficazes na prevenção da doença fatal que é o desfecho do estudo. Tais medidas devem incluir aconselhamento, preservativos e encaminhamento para programas sobre abuso de substâncias e de troca de agulhas. Os pesquisadores também devem adotar essas cointervenções para evitar danos aos participantes no ensaio clínico subsequente sobre a vacina, mesmo que o poder do estudo seja reduzido.

Capítulo 15 Elaborando questionários, entrevistas e pesquisas *on-line*

1a. Não há definição para o tamanho de um *drink*.

1b. Não há como responder se o sujeito toma mais de 8 *drinks* por dia.

1c. A questão não especifica o tempo – finais de semana *versus* dias de semana, todos os dias *versus* menos do que diariamente.

1d. Pode ser melhor especificar um intervalo de tempo (p. ex., nos

últimos 7 dias).

2a. Qual das seguintes alternativas melhor descreve a frequência com que você ingeriu bebida alcoólica ao longo do último ano? Uma bebida alcoólica inclui vinho, destilados ou coquetéis. Selecione uma das oito categorias a seguir:

- ○ Todos os dias
- ○ 5-6 dias por semana
- ○ 3-4 dias por semana
- ○ 1-2 dias por semana
- ○ 2-3 vezes por mês
- ○ Aproximadamente + vez por mês
- ○ Menos de 12 vezes por ano
- ○ Raramente ou nunca

2b. Ao longo do último ano, quantas doses você geralmente tomou em um dia típico em que você bebeu álcool? Considere como dose 350 mL de cerveja, 150 mL de vinho ou 45 mL de destilados. ___________ doses

2c. Ao longo do último ano, qual o maior número de doses de bebida alcoólica que você lembra ter tomado em um dia? ___________ doses

2d. Qual a sua idade quando você começou a beber álcool? ___________ anos. (Se você nunca consumiu bebida alcoólica, escreva “nunca”.)

2e. Já houve algum período em que você bebeu muito mais do que bebe atualmente?

○ Sim →
○ Não ↓

Se Sim, qual das alternativas a seguir melhor descreve a frequência com que você bebia álcool durante aquele período? Selecione uma das 8 categorias.

e(i).
- ○ Todos os dias
- ○ 5-6 dias por semana
- ○ 3-4 dias por semana
- ○ 1-2 dias por semana
- ○ 2-3 vezes por mês
- ○ Aproximadamente 1 vez por mês
- ○ Menos de 12 vezes por ano
- ○ Raramente ou nunca

e(ii). Durante aquele período, quantas doses você geralmente tomava em um dia típico em que bebia álcool? ______ doses

e(iii). Durante quantos anos você bebeu mais do que bebe agora? ______ anos

2f. Você já teve problemas relacionados ao álcool?

Sim

Não

3a. Obter dados através de entrevistas exige muito mais treino e toma muito mais tempo do pessoal de pesquisa do que um questionário de autopreenchimento e, portanto, seu custo é muito maior.
3b. Certos indivíduos não gostam de responder a outra pessoa perguntas delicadas sobre comportamento sexual.
3c. A não ser que os entrevistadores estejam bem treinados e que as entrevistas sejam padronizadas, a informação obtida pode ter grande variabilidade.
3d. No entanto, os entrevistadores podem repetir a questão e orientar o participante de forma a melhorar a compreensão e produzir respostas mais acuradas e completas do que as de um questionário de autopreenchimento.

Capítulo 16 Gerenciamento de dados

1.

IDSujeito	JaFumou100Cigs	IdaPrimCig	MedCigsDia	AlgumCigUltSem	CigsDiaUltSem	IdaParFum
1001	1	17	30	1	30	
1002	1	21	20	0		45
1003	0			0		

Esta tabela mostra como os dados apareceriam em um programa de planilha eletrônica como o Excel. Existem muitas possibilidades aceitáveis para nomes dos campos (cabeçalhos das colunas). Nesta planilha, usam-se iniciais maiúsculas no meio do nome para separar suas partes (IntraCaps). Os especialistas no delineamento de bancos de dados estão igualmente divididos entre aqueles que gostam de usar IntraCaps e aqueles que detestam.

2a.

IdSujeito	NumKit	DataAdm	HrAdm	FimConv-PreHosp	HrFimConv-PreHosp	HrChegHosp	AtivConv-ChegHosp	GCSV-ChegHosp
189	A322	3/12/1994	17:39	0		17:48	1	
410	B536	12/1/1998	01:35	1	01:39	01:53	0	4

2b.

Nome do campo	Tipo de dado	Descrição	Regra de validação
IdSujeito	Número inteiro	Identificador único do sujeito	
NumKit	Texto(5)	Código da farmácia de = caracteres	
DataAdm	Data	Data da administração do fármaco do estudo	
HrAdm	Hora	Hora da administração do fármaco do estudo	
FimConvPreHosp	Sim/não	A convulsão terminou antes da chegada ao hospital?	
HrFimConvPreHosp	Hora	Hora em que a convulsão terminou antes da chegada ao hospital (em branco se a convulsão não terminou)	
HrChegHosp	Hora	Hora da chegada ao hospital	
AtivConvChegHosp	Sim/não	Houve continuação da atividade convulsiva após a chegada ao hospital?	Verificar com FimConvPreHosp
GCSVChegHosp	Número inteiro	Componente verbal da escala de Glasgow na chegada ao hospital (em branco se a convulsão continuou)	Entre 1 e 5

2c. **Vantagens do formulário na tela do computador:**

- Não é necessário transcrever a partir de formulários em papel para tabelas de dados no computador
- *Feedback* imediato sobre entradas inválidas
- Lógica de saltos programada (se a convulsão terminou durante a fase pré-hospitalar, o formulário pergunta a hora em que a convulsão terminou; caso contrário, esse campo é desabilitado e saltado)
- Pode ser disponibilizado por meio de um navegador da internet em vários centros simultaneamente

Desvantagens do formulário na tela do computador:

- Exigências de equipamentos – uma estação de computador
- Exigências de treinamento

Vantagens do formulário em papel:

- Facilidade e velocidade de uso
- Portabilidade
- Permite inserir informações não antecipadas ou dados não estruturados (observações nas margens, respostas que não haviam sido consideradas, etc.)
- Equipamento necessário: uma caneta
- Treinamento já recebido por todo o pessoal do estudo no ensino fundamental

Desvantagens do formulário em papel:

- Exige transcrição subsequente no banco de dados computadorizado
- Sem *feedback* interativo ou lógica de saltos automatizada
- Visualização dos dados e entrada limitados a uma pessoa em um lugar

Embora a entrada de dados por meio de formulários na tela do computador tenha muitas vantagens e a recomendemos para a maioria dos estudos, nesse estudo ela se torna impraticável. A forma mais simples, rápida e fácil de capturar dados em um meio não volátil ainda é usar caneta e papel.

3. Quando codificado como 0 para não ou ausente e + para sim ou presente, o valor médio de uma variável dicotômica (sim/não) é interpretável como a proporção com o atributo. Dentre os participantes randomizados para receber lorazepam, 40,9% (27 de 66) ainda estavam convulsionando quando chegaram ao hospital; daqueles randomizados para receber diazepam, 57,4% (39 de 68) ainda estavam convulsionando; e daqueles randomizados para receber placebo, 78,9% (56 de 71) ainda estavam convulsionando.

Capítulo 17 Implementação do estudo e controle de qualidade

1a. Nada que ele faça será suficiente! Mas alguns passos que ele pode

tomar são listados a seguir:

- Identificar todos os valores faltantes e extremos e rechecar os formulários para garantir que todos os dados foram inseridos corretamente.
- Recuperar dados faltantes de prontuários.
- Coletar dados faltantes de entrevistas a partir dos pacientes que sobreviveram (mas isso não funciona para aqueles que morreram ou para aqueles cujas respostas possam ter mudado durante o período de seguimento).
- Fazer um esforço concentrado para localizar sujeitos perdidos no seguimento, e, pelo menos, obter deles uma entrevista por telefone.
- Buscar o estado vital por meio do National Death Index ou de uma empresa que oferece serviços de localização.

1b. Coletar menos dados.

- Checar os formulários no local imediatamente após a coleta dos dados para garantir que todos os itens estejam completos e corretos.
- Usar entrada interativa de dados com verificações para dados faltantes e fora de faixa.
- Fazer isso logo após a coleta de dados, de modo que os dados faltantes possam ser coletados antes de o paciente deixar o hospital (ou morrer).
- Tabular periodicamente as distribuições de valores para todos os itens durante o curso do estudo para identificar valores faltantes, extremos e potenciais erros.
- Realizar reuniões periódicas com a equipe para avaliar o progresso e enfatizar a importância de dados completos.

Capítulo 18 Estudos comunitários e internacionais

1a. Ambulatório de gastrenterologia

- ***Vantagens:*** É uma provável fonte conveniente e acessível de pacientes. O pessoal do ambulatório provavelmente tem experiência em pesquisa. Não deve ser difícil implementar uma bateria-padrão de testes diagnósticos para pacientes com dor abdominal.
- ***Desvantagens:*** Os pacientes nesse ambulatório podem constituir uma amostra muito selecionada da totalidade dos pacientes com dor abdominal na comunidade, e o curso clínico desses pacientes pode ser diferente de outros na comunidade. Assim, os resultados podem ter baixa capacidade de generalização.

1b. Unidades de atenção primária à saúde

- ***Vantagens:*** É possível identificar pacientes na primeira manifestação do quadro clínico, sem a seleção e o atraso causados pelo processo de referência. Médicos comunitários podem ser beneficiados pela oportunidade de participar em pesquisa.
- ***Desvantagens:*** São principalmente de ordem logística. Identificar médicos e pacientes e implementar um protocolo de pesquisa-padrão pode demandar um grande esforço de organização, e será um desafio garantir o controle de qualidade.

2a. Essa questão só pode ser respondida com dados locais. Pesquisas em outros lugares não se aplicam.

2b. Essa questão já está amplamente discutida na literatura internacional. Repetir tais pesquisas na China não é uma forma eficiente de aproveitar os recursos.

2c. Para essa questão, a capacidade de generalização de pesquisas de outros lugares provavelmente é intermediária. Estratégias para parar de fumar que se mostraram bem-sucedidas em outros países podem servir como base para estratégias a serem tentadas na China. No entanto, sem pesquisa local, não há como garantir que elas também serão bem-sucedidas na população chinesa. Estudos anteriores em populações de outros lugares com laços culturais fortes com a China, como chineses que migraram recentemente para os Estados Unidos, podem ser particularmente úteis.

Capítulo 19 Redigindo uma proposta para solicitar financiamento de pesquisa

1-3. Esperamos que você tenha algumas ideias úteis para planejar sua própria agenda de pesquisa, e o encorajamos a envolver seus mentores e pares em discussões sobre a melhor forma de prosseguir.

Glossário

Abandono. Um sujeito do estudo para o qual não é possível determinar o desfecho, muitas vezes porque recusaram o seguimento. Às vezes, isso inclui os sujeitos que abandonaram porque foram a óbito durante o estudo. Por exemplo, houve 17 abandonos em um estudo: oito devido à recusa e três porque desenvolveram demência.

Acurácia. É o quanto uma medida corresponde ao seu valor verdadeiro. Por exemplo, o peso autorreferido é uma medida menos acurada do peso real do que uma medida feita utilizando uma balança calibrada.

Ajuste multivariado. Termo genérico para as técnicas estatísticas utilizadas para ajustar para os efeitos de uma ou mais potenciais variáveis confundidoras sobre a associação entre um preditor e um desfecho. Por exemplo, utilizando ajuste multivariado, um estudo mostrou que a ingesta de suplementos de vitamina D estava associada com risco aumentado de declínio cognitivo, ajustado para idade, sexo, escolaridade, função cognitiva basal e tabagismo.

Ajuste. É um termo genérico para uma série de técnicas estatísticas utilizadas para controlar os efeitos de uma ou mais variáveis em uma associação entre duas outras variáveis. Por exemplo, o ajuste para renda reduziu a magnitude da associação entre escolaridade e mortalidade.

Alfa. Ao delinear um estudo, corresponde ao valor máximo pré-estabelecido para a probabilidade de incorrer em um erro Tipo I, isto é, rejeitar a hipótese nula quando ela é verdadeira. Por exemplo, ao selecionar um alfa de 0,05, o investigador determinou uma probabilidade máxima de 5% de que o seu estudo irá encontrar uma associação estatisticamente significativa entre raça não branca e risco de câncer de cólon tão somente pelo acaso. Também denominado nível de significância estatística.

Amostra aleatória. Amostra sorteada enumerando as unidades da população e selecionando um subconjunto aleatoriamente. Por exemplo, uma amostra aleatória de pessoas com catarata na clínica do investigador exigiria que o investigador listasse todos os pacientes com catarata e usasse uma tabela de números aleatórios ou números aleatórios gerados pelo computador para selecionar a amostra. Ver também *amostra probabilística*.

Amostra consecutiva. Uma amostra de estudo na qual os sujeitos são selecionados consecutivamente até atingir o tamanho de amostra desejado. Esse termo, em geral, se refere à amostra pretendida. Pode referir-se também à amostra efetivamente estudada quando se faz revisão de registros médicos, pois o consentimento informado pode não ser necessário nesse caso. Por exemplo, os investigadores realizaram amostragem consecutiva para revisar os prontuários dos primeiros 100 pacientes com artrite reumatoide vistos na clínica de reumatologia, iniciando em 15 de janeiro de 2013.

Amostra de base populacional. Amostra de pessoas que representam toda uma população. Por exemplo, o National Health and Nutrition Examination Survey (NHANES), que fornece dados sobre uma amostra aleatória de toda a população dos Estados Unidos, é uma amostra de base populacional.

Amostra de conveniência. Um grupo de sujeitos que foram selecionados para um estudo apenas porque eram de acesso relativamente fácil. Por exemplo, o investigador utilizou uma amostra de

conveniência de pacientes de seu ambulatório para servirem como controles no seu estudo de caso-controle sobre fatores de risco para meningioma.

Amostra pretendida. Grupo de sujeitos que o investigador pretende incluir em um estudo, como descrito no protocolo. Por exemplo, a amostra pretendida para um estudo consistia em mulheres com câncer de mama que foram vistas inicialmente para tratamento em uma segunda ou quinta-feira no Longview Hospital (dias em que o investigador ou sua equipe de pesquisa estavam disponíveis) e que tinham recebido o diagnóstico menos de seis semanas antes, durante o período entre 1º de janeiro de 2013 e 30 de junho de 2014. Ver também *população acessível* e *amostra*.

Amostra representativa. Amostra de pessoas arroladas em um estudo que é representativa da população-alvo. Por exemplo, no Framingham Heart Study, a população-alvo era todos os adultos. A população acessível (aos investigadores localizados em Boston) era a população adulta na cidade de Framingham, Massachussetts. Os investigadores enumeraram os adultos de Framingham, e, a cada dois adultos listados, um era convidado a participar do estudo. Essa abordagem poderia resultar em uma amostra representativa, mas algumas pessoas se recusaram a participar e foram substituídas por voluntários. Como os voluntários muitas vezes têm hábitos de vida mais saudáveis que os não-voluntários, a amostra pode ter hiper-representado pessoas saudáveis. Além disso, a população de Framingham (que era majoritariamente branca) não representa a população norte-americana e, certamente, não representa os adultos de outros países.

Amostra sistemática. Amostra que é sorteada enumerando as unidades da população elegível e selecionando um subconjunto da população por meio de um processo pré-especificado. Por exemplo, no Framingham Heart Study, os investigadores construíram uma lista de todos os adultos residentes na cidade de Framingham, Massachusetts, e então selecionaram um a cada dois residentes para serem incluídos no estudo como parte da amostra sistemática.

Amostra. Subconjunto da população que participa de um estudo. Por exemplo, em um estudo sobre um novo tratamento para a asma, onde a população-alvo é todas as crianças com asma e a população acessível é as crianças com asma na cidade do investigador neste ano, a amostra do estudo é as crianças na cidade do investigador neste ano que de fato foram inscritas no estudo.

Amostragem aleatória estratificada. Técnica de amostragem na qual os potenciais participantes são estratificados em grupos com base em características, como idade, raça ou sexo, e uma amostra aleatória é sorteada dentro de cada estrato. Os estratos podem ser ponderados de diversas formas. Por exemplo, os investigadores usaram a amostragem aleatória estratificada em um estudo sobre a prevalência de câncer de pâncreas na Califórnia para hiperamostrar minorias raciais e étnicas.

Amostragem por conglomerados. Uma técnica de amostragem na qual os sujeitos são selecionados em grupos (conglomerados), em vez de como indivíduos. Geralmente é utilizada devido à conveniência ao amostrar grandes populações. Por exemplo, um investigador interessado em determinar a prevalência de uso de medicamentos utilizou amostragem por conglomerados para arrolar 300 pacientes. Primeiro ele identificou potenciais sujeitos escolhendo 10 prefixos de três dígitos (p.ex., 285-, 336-, etc.) dentro de um código de área. Então, utilizou a discagem de dígito aleatório para encontrar 30 sujeitos dispostos a participar dentro de cada conglomerado com o mesmo prefixo de três dígitos.

Amostragem por incidência-densidade. Em um estudo de caso-controle aninhado, é uma técnica para selecionar controles quando uma exposição importante muda com o tempo; assim, a exposição precisa ser medida em um momento semelhante nos casos e nos controles. Por exemplo, um estudo de caso-controle aninhado para determinar se o uso de anti-histamínicos, que tem variação sazonal aumenta o risco a curto prazo de fraturas de quadril (presumivelmente devido a um aumento no risco de queda) usou a amostragem por incidência-densidade dos controles, de modo que o uso de anti-histamínico pelo

controle foi medido durante o mesmo mês que uma fratura de quadril ocorreu em um caso.

Amostragem probabilística. Processo aleatório, geralmente realizado utilizando uma tabela de número aleatórios ou um algoritmo de computador, para garantir que cada membro de uma população tenha uma probabilidade especificada de ser incluído na amostra, fornecendo, assim, uma base rigorosa para permitir inferências a partir da amostra para a população. Por exemplo, uma observação de uma amostra probabilística de 5% das pessoas com doença pulmonar obstrutiva crônica (DPOC) baseada nos diagnósticos da alta hospitalar de todos os hospitais da Califórnia deveria fornecer achados confiáveis sobre fatores de risco para reinternação e morte.

Amostragem. Processo de selecionar participantes para um estudo quando o número de participantes elegíveis é maior do que o tamanho de amostra estimado. Por exemplo, o investigador utilizou um esquema de amostragem "1 em 3" para selecionar, em média, um de cada três sujeitos elegíveis. Ver também *amostragem por conglomerados*, *amostra consecutiva*, *amostra de conveniência*, *amostra probabilística*, *amostragem aleatória estratificada* e *amostra sistemática*.

Análise de dados secundários. Uso de dados existentes para investigar questões de pesquisa que não são aquelas para as quais os dados foram originalmente coletados. Bancos de dados secundários podem incluir estudos prévios, registros médicos, dados de cobrança de seguros de saúde e declarações de óbito. Por exemplo, os dados de altas hospitalares e os registros de declarações de óbito podem ser usados em uma análise de dados secundários para determinar a mortalidade, em 1 ano, de pacientes com diagnóstico de pancreatite aguda na alta.

Análise de sensibilidade. Utilização de diferentes métodos (p. ex., definições alternativas das variáveis preditoras ou de desfecho, diferentes testes estatísticos) para determinar se os resultados da análise principal são robustos. Por exemplo, em uma metanálise de ensaios clínicos sobre o efeito de inibidores seletivos da receptação da serotonina sobre a depressão, em uma análise de sensibilidade, o investigador poderia incluir apenas os ensaios clínicos cegados para demonstrar que os resultados são robustos quando a análise é restrita aos estudos com boa qualidade metodológica.

Análise de sobrevivência. Técnica estatística utilizada para comparar os tempos até um desfecho entre os grupos de um estudo, não necessariamente a sobrevivência. Por exemplo, em um ensaio clínico randomizado sobre o efeito da cirurgia de revascularização miocárdica quando comparada com a angioplastia percutânea para a prevenção de infarto do miocárdio e morte, poderia-se utilizar a análise de sobrevivência para comparar o tempo desde o início do tratamento até a ocorrência desses desfechos nos dois grupos.

Análise de subgrupos. Comparações entre grupos randomizados em um subconjunto dos participantes do ensaio clínico. Por exemplo, em um ensaio clínico randomizado sobre o efeito do modulador seletivo do receptor estrogênico (MSRE) sobre a recorrência de câncer de mama, os investigadores realizaram uma análise de subgrupos sobre o efeito do tratamento de acordo com o estádio do câncer, comparando o efeito do MSRE contra o placebo em mulheres com estádios I, II, III e IV da doença.

Análise por intenção de tratar. Em um ensaio clínico randomizado, é o processo de comparar sujeitos com base no grupo ao qual foram randomizados, mesmo se ele não corresponder ao tratamento que o sujeito de fato recebeu. Essa é a forma mais rigorosa de análise. Por exemplo, os investigadores realizaram uma análise por intenção de tratar para determinar se a alocação aleatória para receber seis meses de psicoterapia melhorava sintomas de ansiedade quando comparada com a alocação aleatória a um grupo-controle que recebia um panfleto sobre redução de estresse. Ver também *análise conforme o protocolo*.

Análise por protocolo. Em ensaio clínico, é uma abordagem de análise em que os dados dos participantes são apenas incluídos se estes participantes aderiram ao protocolo do estudo, o que é em

geral definido como ter realizado a intervenção do estudo conforme foram orientados. Por exemplo, em um ensaio clínico randomizado sobre cirurgia comparada com fisioterapia para o tratamento da artrose grave do joelho, uma análise por protocolo incluiria dados apenas de participantes no grupo de cirurgia que realmente realizaram a cirurgia e de participantes no grupo de fisioterapia que aderiram ao plano de fisioterapia. Ver também *análise por intenção de tratar*.

Associação espúria. Associação entre uma variável preditora e uma variável de desfecho que é vista em um estudo mas não é verdadeira na população, seja devido ao acaso ou a viés. Por exemplo, estudos observacionais mostraram uma redução no risco de doença cardiovascular em pessoas que tomavam suplementos de betacaroteno. Entretanto, um ensaio clínico randomizado sobre suplementos de beta caroteno não mostrou efeito sobre o risco de doença cardiovascular, sugerindo que a associação observada nos estudos observacionais era espúria.

Associação. Uma relação quantificável entre duas variáveis. Por exemplo, o estudo encontrou uma associação entre o sexo masculino e o risco de declínio cognitivo em pessoas com idade entre 60 e 69 anos, com uma razão de riscos de 1,6.

Banco de dados relacional. *Software* que permite armazenar informações relacionadas em uma série de tabelas. As tabelas podem ser relacionadas umas com as outras por meio de campos em comum. Por exemplo, um banco de dados relacional para um estudo poderia incluir os campos IDSujeito e DataNasc em uma tabela de Sujeitos e IDSujeito e DataVisita em uma tabela de Encontros, que poderia ter muitos encontros por sujeito. A idade de um participante na data do encontro pode ser calculada facilmente utilizando o IDSujeito para relacionar cada DataVisita com a data de nascimento daquele participante.

Beta. Ao delinear um estudo, a probabilidade máxima pré-estabelecida de incorrer em um erro Tipo II, isto é, deixar de rejeitar a hipótese nula quando ela é falsa. Essa medida faz sentido apenas no contexto da magnitude de efeito. Por exemplo, se um investigador especifica um beta de 0,20 (e um alfa de 0,05), seriam necessários aproximadamente 25 mil sujeitos por grupo, acompanhados por 10 anos, para mostrar que o uso diário de aspirina diminui pela metade o risco de câncer de cólon. Se a aspirina realmente tiver esse efeito, o seu estudo com 25 mil pessoas por grupo teria uma probabilidade de 20% de deixar de rejeitar a hipótese nula de ausência de diferença (a um alfa = 0,05). Ver também *poder estatístico*.

Calibração. Processo de assegurar que um instrumento forneça leituras consistentes. A calibração geralmente é feita aferindo um padrão conhecido e então ajustando (calibrando) o instrumento de acordo com esse padrão. Por exemplo, a balança era calibrada mensalmente utilizando um bloco de aço de 50 kg.

Campo. Uma coluna em uma tabela de um banco de dados relacional que inclui dados sobre um atributo específico do registro. Por exemplo, dois campos na tabela de Encontros foram IdSujeito (para relacionar com as informações específicas dos sujeitos) e PesoKg (peso em kg).

Capacidade de generalização. É o quanto os resultados de uma amostra de um estudo se aplicam a outras populações. Por exemplo, o revisor questionou a capacidade de generalização da taxa de sucesso de 90% relatada para a radioablação intraluminal de membrnas do esôfago inferior, uma vez que os procedimentos foram todos realizados pelo gastrenterologista que desenvolveu e aperfeiçoou a técnica em 350 pacientes, ao passo que a maioria dos gastrenterologistas só irá ver alguns poucos pacientes com esse mesmo problema ao longo de suas carreiras.

Caso. Sujeito que tem ou desenvolve o desfecho de interesse. Por exemplo, os casos foram definidos como aquelas pessoas que desenvolveram angina instável, infarto do miocárdio ou morte súbita durante o seguimento. Ver também *controle*.

Causa-efeito. Conceito de que um preditor é responsável por produzir um desfecho – ou aumenta a probabilidade de sua ocorrência. O objetivo da maioria dos estudos observacionais é demonstrar causa-efeito, embora isso seja difícil de fazer, a não ser que a causa (p.ex., um tratamento) seja alocada aleatoriamente. Por exemplo, o investigador realizou um estudo de caso-controle para determinar se havia uma relação de causa-efeito entre consumo de álcool (a causa) e câncer de pâncreas (o efeito). Ver também *confundimento* e *efeito-causa*.

Cegamento. Processo de assegurar que os sujeitos e/ou investigadores não estejam cientes do grupo (p.ex., intervenção ou controle) em que os sujeitos foram alocados, geralmente no contexto de um ensaio clínico randomizado. Também denominado mascaramento, especialmente em estudos sobre oftalmologia. Por exemplo, ao usar comprimidos idênticos placebo e manter a lista das alocações dos sujeitos guardada em local fora do espaço onde foi realizado o estudo, tanto os sujeitos quanto os investigadores (incluindo os assistentes da pesquisa) foram cegados quanto aos sujeitos que receberam o medicamento ativo.

Chances. Risco de uma doença (ou de outro desfecho) dividido por 1 – risco. Por exemplo, se o risco ao longo da vida de desenvolver câncer de mama em mulheres é de 15%, então as chances ao longo da vida de desenvolver câncer de mama são de 0,18 (0,15/0,85). O risco e as chances são semelhantes para doenças raras (aquelas que ocorrem em menos de 10% das pessoas).

Chave principal. Em um banco de dados relacionais, é o campo ou combinação de campos que identificam de forma única cada coluna em uma determinada tabela. Por exemplo, o investigador criou um campo NumeroVisita único para servir como chave principal de uma tabela de consultas ambulatoriais.

***Classification and Regression Trees* (CART)**. Ver *particionamento recursivo*.

Coeficiente de correlação. Termo estatístico que indica o grau em que duas medidas contínuas estão relacionadas linearmente, de modo que uma mudança em uma das medidas está associada com uma mudança proporcional na outra. Frequentemente abreviado como *r*. Por exemplo, a altura e o peso foram correlacionados em uma amostra de mulheres de meia idade, com um $r = 0{,}7$.

Coeficiente de variação (CV). Medida da precisão de uma aferição, obtida dividindo o desvio-padrão de uma série de aferições realizadas em uma única amostra pela média dessas aferições. Às vezes, o CV é obtido para valores no meio e nos extremos da aferição. Por exemplo, o laboratório determinou que seu coeficiente de variação para os níveis de estradiol sérico era de 10% em uma amostra de mulheres na perimenopausa (nas quais o estradiol estava muito baixo), mas de apenas 2% em mulheres mais jovens.

Cointervenção. Em um ensaio clínico, é uma intervenção que ocorre após a randomização, mas que não seja a intervenção em estudo, e que afeta a probabilidade de um desfecho. Cointervenções que ocorrem em diferentes taxas nos grupos do estudo podem enviesar o desfecho e tornar difícil afirmar que esse desfecho foi causado pela intervenção em estudo. Por exemplo, um estudo sobre o efeito de uma intervenção de promoção do aleitamento materno sobre doenças alérgicas no futuro em lactentes foi difícil de interpretar, pois as mulheres no grupo de intervenção não apenas amamentaram por mais tempo, mas também tiveram maior probabilidade do que o grupo-controle em retardar a introdução de alimentos sólidos e de comprar fórmulas hipoalergênicas, ambos os quais foram potenciais cointervenções.

Comitê diretivo. Em um estudo multicêntrico, é um comitê que assegura a governança geral para o estudo. Geralmente é composto pelos investigadores principais de cada centro do estudo, do centro de coordenação e representantes da instituição financiadora. Por exemplo, o comitê diretivo do estudo decidia se os estudos suplementares propostos deveriam ser realizados.

Concordância. Uma medida de concordância entre dois (ou mais) observadores sobre a ocorrência de um fenômeno. Por exemplo, a concordância entre os radiologistas A e B foi de 96% para a presença de um infiltrado lobar na radiografia de tórax, mas de apenas 76% para a presença de cardiomegalia. Ver também *kapa*.

Condicionamento em um efeito comum. Uma fonte de viés em estudos epidemiológicos nos quais é introduzida uma associação entre duas diferentes causas do mesmo efeito por meio do condicionamento naquele efeito. Por exemplo, devido ao condicionamento em um efeito comum (tempo total de televisão ligada) há uma associação inversa entre assistir à televisão e jogar videogame em crianças que ficam com a televisão ligada pelo menos seis horas por dia. Aquelas que passam mais tempo assistindo à televisão jogam menos videogame.

Condicionamento. Processo de examinar as associações entre duas ou mais variáveis em níveis fixos de outra variável na qual elas estão "condicionadas". A especificação, o pareamento, a estratificação e o ajuste multivariado são as formas mais comuns de condicionamento em uma variável. Por exemplo, o investigador não encontrou associação entre uso de cocaína e risco de sífilis após condicionar o número de parceiros sexuais.

Confundidor. Ver *confundimento*.

Confundimento por indicação. Uma forma específica de confundimento na qual uma das indicações para um tratamento é o confundidor. Em geral, ocorre em estudos observacionais sobre a associação entre um tratamento e um desfecho. Por exemplo, os revisores de um estudo observacional estavam preocupados que a associação relatada entre um novo tratamento para o transtorno bipolar e o aumento do risco de suicídio poderia ter ocorrido porque os pacientes com doença subjacente mais grave haviam propositadamente recebido o novo medicamento.

Confundimento. Um fenômeno epidemiológico no qual uma associação entre uma variável preditora e uma variável de desfecho se dá devido a uma terceira variável (denominada confundidora, ou variável de confusão), e não devido a uma relação de causa-efeito entre a variável preditora e o desfecho. Por exemplo, a associação aparente entre tabagismo e câncer cervical foi confundida pela infecção pelo papilomavírus humano (HPV), uma vez que as mulheres que fumavam tinham maior probabilidade de terem (múltiplos parceiros sexuais e) infecção pelo HPV. Ver também *modificação de efeito*.

Consulta (*query*). Comando ou instrução para um banco de dados relacional para selecionar ou manipular os dados. Por exemplo, o coordenador do estudo fez uma consulta para selecionar nomes e informações de contato para todos os participantes do estudo que estavam previstos para realizarem uma visita de seguimento nos 2 meses subsequentes e que ainda não haviam sido agendados.

Contaminação. O processo indesejável no qual algum ou a maioria dos efeitos de uma intervenção também afetam sujeitos no grupo-controle. Por exemplo, um estudo sobre se os efeitos de ensinar crianças a contar de trás para frente melhorou suas habilidades aritméticas sofreu contaminação, pois as crianças no grupo de intervenção não conseguiam resistir à tentação de ensinar essa habilidade a seus amigos no grupo-controle.

Controle de qualidade. Processos para assegurar que a condução de um estudo, incluindo o arrolamento, as aferições, os procedimentos laboratoriais e o gerenciamento e a análise dos dados, sejam da melhor qualidade possível. Por exemplo, os investigadores controlaram a qualidade da coleta de dados elaborando procedimentos explícitos descritos por escrito para todas as aferições em um manual de operações e observando periodicamente a equipe do estudo para assegurar que ela seguia esses procedimentos.

Controle. Termo com dois significados distintos. No primeiro deles, controle se refere a um sujeito que não desenvolveu o desfecho de interesse e, portanto, é membro de um grupo de comparação com o qual

aqueles que desenvolveram o desfecho (os “casos”) serão comparados. Por exemplo, em um estudo sobre fatores de risco para úlcera péptica, os controles foram selecionados a partir de pacientes hospitalizados durante o período do estudo com diagnóstico de doença não digestiva. No segundo significado, controle se refere ao “tratamento” não ativo (isto é, um placebo ou o cuidado usual) recebido pelos participantes em um ensaio clínico que não receberam a intervenção em estudo. Nesse contexto, o termo controle também é utilizado para designar um participante que recebeu o tratamento não ativo. Por exemplo, os controles receberam comprimidos de placebo com aparência idêntica à dos medicamentos ativos. Ver também *caso* e *intervenção*.

Controle-placebo. Controle inativo que é indistinguível em relação ao medicamento ou intervenção de ativos usados em um ensaio clínico randomizado. Por exemplo, em um ensaio clínico randomizado controlado por placebo sobre um novo tratamento para incontinência urinária, o placebo deve ter aparência, odor, sabor e sensação tátil semelhante ao novo medicamento que está sendo testado.

Controles ambulatoriais. No contexto de um estudo de caso-controle, é a seleção dos controles a partir dos mesmos ambulatórios de onde os casos foram selecionados. Por exemplo, uma investigadora utilizou controles ambulatoriais no seu estudo sobre se correr no asfalto pelo menos duas milhas por semana estava associado a artrose radiográfica do joelho.

Controles hospitalares. No contexto de um estudo de caso-controle, é a seleção dos controles a partir do mesmo hospital de onde foram selecionados os casos. Por exemplo, no seu estudo sobre se comer carne processada estava associado com câncer do trato digestivo superior, uma investigadora usou controles hospitalares selecionados a partir dos pacientes que haviam tido doenças digestivas não malignas tratadas no mesmo hospital que os casos.

Correção de Bonferroni. Uma técnica para prevenir erros Tipo I na qual se divide o alfa global de um estudo pelo número de hipóteses testadas. Por exemplo, uma vez que os investigadores estavam testando quatro diferentes hipóteses, eles usaram a correção de Bonferroni para reduzir o alfa de cada hipótese de 0,05 para 0,0125.

Cox. Ver *modelo de Cox*.

Critérios de entrada. Uma lista de atributos que os sujeitos devem ter para serem elegíveis para participar em um estudo. Os critérios de entrada podem variar se os sujeitos são arrolados em diferentes grupos, como em um estudo de caso-controle, ou em um estudo de dupla coorte. Por exemplo, os critérios de entrada para um estudo sobre um novo tratamento para gota incluíram idade entre 20 a 75 anos, pelo menos um episódio de gota diagnosticada pelo médico nos últimos 12 meses e nível sérico de ácido úrico de pelo menos 6 mg/dL. Ver também *critérios de exclusão* e *critérios de exclusão*.

Critérios de exclusão. Uma lista de atributos que impedem que um potencial sujeito seja elegível para um estudo. Por exemplo, os critérios de exclusão para um estudo eram tratamento prévio com antidepressivo nos últimos dois anos, uso atual de alfa-bloqueadores ou beta-bloqueadores e incapacidade de ler inglês no nível esperado para a 6ª série. Ver também *critérios de inclusão*.

Critérios de inclusão. Lista de atributos exigidos dos potenciais sujeitos de um estudo. Por exemplo, os critérios de inclusão de um estudo incluíam pessoas entre 18 e 65 anos de idade que moravam em São Francisco e não tinham história prévia de depressão. Ver também *critérios de exclusão*.

Critérios de seleção. Regras que definem quem é elegível para arrolamento em um estudo, incluindo os critérios de inclusão e de exclusão. Por exemplo, em um ensaio clínico sobre testosterona transdérmica para melhorar a libido em mulheres pós-menopáusicas, os critérios de seleção poderiam ser mulheres com idade entre 45 e 60 anos com baixa libido, livres de doença arterial coronariana e que não haviam tido mais de três períodos menstruais no último ano.

Cruzamento. Termo utilizado para descrever quando um sujeito, geralmente durante um ensaio clínico, inicia em um grupo (p. ex., cuidado usual) e troca para o outro grupo (p. ex., tratamento ativo). Seu uso mais comum é quando o tratamento ativo envolve um procedimento. Por exemplo, 15 sujeitos com câncer de próstata que haviam sido alocados inicialmente para monitoramento expectante foram cruzados para receber radioterapia ou cirurgia. Ver também *estudo cruzado*.

Curva ROC (característica operatória do receptor). Técnica gráfica para quantificar a acurácia de um teste diagnóstico e ilustrar o contrabalanço entre a sensibilidade e a especificidade em diferentes pontos de corte para definir o teste como positivo. A curva mostra as taxas de verdadeiros positivos (sensibilidade) no eixo Y e as taxas correspondentes de falsos positivos (1 – especificidade) no eixo X em diversos pontos de corte para considerar o teste como positivo. A área sob a curva ROC, que varia de 0,5 para um teste inútil a 1,0 para um teste perfeito, é um resumo útil da acurácia global do teste. Por exemplo, a área sob a curva ROC para o uso da tomografia no diagnóstico da apendicite (que pode ser interpretada como Claramente positiva, Provavelmente positiva, Pouco útil, Provavelmente normal ou Claramente normal) foi de 0,95, um valor substancialmente superior ao valor de 0,77 para a ultrassonografia (que tinha categorias semelhantes Δ einterpretação).

Dados faltantes. Dados que não foram coletados durante um estudo, seja na linha de base ou durante o seguimento. Por exemplo, uma pesquisadora estava preocupada que a proporção relativamente grande (34%) de sujeitos com dados faltantes em relação ao consumo de álcool poderia ter enviesado seu estudo sobre fatores de risco para quedas.

Dados. Termo utilizado para descrever medidas, geralmente em formato numérico. Por exemplo, os dados sobre a prevalência de diversas doenças são úteis ao tomar decisões sobre a alocação de recursos de saúde.

Delineamento de séries temporais. Delineamento intragrupo no qual as aferições são feitas antes e depois de cada participante (ou de toda uma comunidade) receber uma intervenção. Esse delineamento elimina o confundimento, pois cada participante serve como seu próprio controle. Entretanto, os delineamentos intragrupos são suscetíveis aos efeitos do aprendizado, à regressão à média e a tendências seculares. Por exemplo, por meio de um delineamento de séries temporais, a glicemia de jejum foi aferida em um grupo de pacientes com diabetes antes de iniciar um programa de exercícios físicos e novamente após o programa ter sido concluído, para determinar se o exercício físico reduzia a glicemia de jejum. Ver também *delineamento intragrupo*.

Delineamentos intergrupos. Um delineamento que compara as características ou desfechos de sujeitos em dois (ou mais) grupos diferentes. Por exemplo, o investigador usou um delineamento intergrupos para comparar as taxas de mortalidade hospitalar em pacientes atendidos em unidades de tratamento intensivo com intensivistas em tempo integral com as taxas em pacientes tratados em unidades que utilizavam monitoramento eletrônico dos pacientes. Ver também *delineamento intragrupo*.

Delineamentos intragrupo. Delineamento no qual as aferições são comparadas em um único grupo de participantes, geralmente em dois períodos de tempo diferentes. Esse delineamento elimina o confundimento, pois cada participante serve como seu próprio controle. Entretanto, os delineamentos intragrupos são suscetíveis aos efeitos do aprendizado, à regressão à média e a tendências seculares. Por exemplo, por meio de um delineamento intragrupo, a glicemia de jejum foi aferida em um grupo de pacientes com diabetes antes de iniciar um programa de exercícios físicos e novamente após o programa ter sido concluído, para determinar se o exercício físico reduzia a glicemia de jejum. Ver também *delineamentos intergrupos, teste t para uma amostra* e *delineamento de séries temporais*.

Desfecho. Termo genérico para os eventos finais que podem ocorrer em um estudo, como morte ou ocorrência de uma doença. Por exemplo, em um estudo sobre se a radiocirurgia era benéfica para

pacientes com metástase cerebral solitária, os sujeitos foram seguidos para desfechos de morte ou internação em instituições de longa permanência de idosos.

Desvio-padrão. Medida da variância (dispersão) de uma variável contínua. Por exemplo, o investigador relatou que a idade média na coorte de 450 homens era de 59 anos, com um desvio padrão de 10 anos.

Dicionário de dados. Tabela ou planilha que inclui informações sobre cada uma das variáveis em um estudo, incluindo seu nome e tipo (p. ex., numérica, ou de texto), a definição de cada valor e a faixa permitida de valores. Por exemplo, o investigador consultou o dicionário de dados porque havia esquecido que o número "5" no campo "raça" era usado para indicar Indígena Norte-Americano ou Nativo do Alasca.

Diferença de riscos. Risco de um desfecho em um grupo menos o risco em um grupo de comparação. Por exemplo, se o risco de eventos tromboembólicos em mulheres usuárias atuais de estrogênio for de 5/1000 (0,5%) e o risco naquelas que nunca usaram estrogênio for de 2/1000 (0,2%), a diferença de riscos em mulheres usando estrogênio quando comparadas às não usuárias é de 3/1000 (0,3%). Ver também *número necessário tratar*.

Dose-resposta. O fenômeno segundo o qual quanto maior a exposição (dose), maior é a magnitude ou probabilidade do desfecho (resposta). (Se uma exposição for protetora, então, quanto maior a exposição, menor a probabilidade do desfecho.) Por exemplo, um estudo relatou uma relação dose-resposta entre exposição solar e número de nevos melanocíticos; outro estudo relatou uma relação dose-resposta entre número de nevos e risco de melanoma.

Efeito-causa. Situação na qual um desfecho causa o preditor, e não o contrário. Por exemplo, embora um estudo de caso-controle tenha observado que a exposição a broncodilatadores inalatórios estava associada a um risco aumentado de doença intersticial pulmonar, a explicação mais provável era efeito-causa, uma vez que os pacientes com doença intersticial pulmonar tinham maior probabilidade de terem sido tratados (erroneamente) com inaladores. Ver também *causa-efeito*.

Efeito-sumário. Em uma metanálise, é o efeito médio ponderado baseado nos estudos incluídos; a fórmula para os pesos depende do modelo utilizado. Por exemplo, em uma metanálise de ensaios clínicos randomizados sobre o efeito de um inibidor da enzima conversora da angiotensina (ECA) sobre a mortalidade em pacientes com doença coronariana, o efeito-sumário com o modelo de efeitos fixos foi o risco relativo médio ponderado, sendo utilizado como peso o inverso da variância do risco relativo em cada um dos estudos incluídos. Ver também *modelo de efeitos fixos* e *modelo de efeitos aleatórios*.

Efetividade. Embora não haja definição-padrão para esse termo, costuma-se defini-lo como uma medida de quão bem uma intervenção funciona na prática real, em oposição a quão bem ela funcionou em um ensaio clínico randomizado. Por exemplo, como ensaios clínicos mostraram que o ativador do plasminogênio tecidual (tPA) reduzia a morbimortalidade por acidente vascular encefálico em diversos ensaios clínicos realizados em regiões urbanas, um grupo de investigadores estudou sua efetividade em 25 serviços de emergência de áreas rurais. Ver também *eficácia*.

Eficácia. Embora não haja definição-padrão para esse termo, costuma-se defini-lo como uma medida de quão bem uma intervenção funcionou em um ensaio clínico, em oposição a como funcionaria na prática real. Por exemplo, um ensaio clínico mostrou que o ativador do plasminogênio tecidual (tPA) tinha uma eficácia de 25% na redução da morbimortalidade em pacientes com acidente vascular encefálico. Ver também *efetividade*.

Ensaio clínico de fase I. Ensaio clínico inicial, geralmente não cego e não controlado, testando doses crescentes de um novo tratamento em um número pequeno de voluntários humanos para avaliar sua segurança. Por exemplo, um ensaio clínico de fase I sobre um novo medicamento para tratamento de

fogachos na menopausa geralmente incluiria um pequeno número de voluntárias (com ou sem fogachos) que receberiam doses crescentes do medicamento para determinar seus efeitos nos valores do hemograma e da função hepática e renal, nos achados do exame físico, nos sintomas e em outros eventos adversos inesperados.

Ensaio clínico de fase II. Pequeno ensaio clínico randomizado (de preferência cego) para testar o efeito de uma série de diferentes doses de um novo tratamento sobre efeitos adversos, assim como sobre desfechos substitutos ou clínicos. Por exemplo, um ensaio clínico de fase II sobre um novo medicamento para fogachos que havia sido mostrado como seguro em um ensaio clínico de fase I poderia arrolar um pequeno número de mulheres pós-menopáusicas com fogachos, alocá-las aleatoriamente para duas ou três diferentes doses do novo medicamento ou placebo e então acompanhá-las para determinar a frequência de fogachos, assim como de efeitos adversos.

Ensaio clínico de fase III (estudo central). Ensaio clínico randomizado (de preferência cego) grande o suficiente para testar a eficácia e a segurança de um novo tratamento. Por exemplo, se a melhor dose de um novo tratamento para fogachos tiver sido estabelecida em um ensaio clínico de fase II e o novo tratamento tiver sido seguro, o próximo passo seria um grande ensaio clínico de fase III em que mulheres com fogachos seriam aleatoriamente alocadas para o novo tratamento ou para placebo e acompanhadas para a ocorrência de fogachos e efeitos adversos.

Ensaio clínico de fase IV. Estudo de grande porte, que pode ou não ser um ensaio clínico randomizado, conduzido após um medicamento ter sido aprovado por uma agência regulatória como o US Food and Drug Administration (FDA), muitas vezes para determinar a segurança do medicamento ao longo de um tempo maior do que seria possível em um ensaio clínico de fase III. Por exemplo, após um novo medicamento para o tratamento de fogachos na menopausa ter sido aprovado pelo FDA, um ensaio clínico de fase IV poderia incluir mulheres com fogachos menos intensos do que aquelas incluídas no ensaio clínico de fase III.

Ensaio clínico de não inferioridade. É um ensaio clínico que compara um novo tratamento que tem algumas vantagens sobre um tratamento já estabelecido (p. ex., o novo tratamento é mais seguro, barato ou fácil de utilizar), com o objetivo de demonstrar que o novo tratamento não é inferior ao tratamento estabelecido. Por exemplo, um ensaio clínico sobre um novo analgésico que não causa sonolência demonstrou que o novo medicamento não era inferior à oxicodona para o alívio da dor pós-operatória.

Ensaio clínico fatorial. Um ensaio clínico sobre dois ou mais tratamentos (p. ex., A e B), às vezes com desfechos não relacionados, nos quais os sujeitos são alocados aleatoriamente para receberem o tratamento A ativo e o placebo B, o tratamento ativo B e o placebo A, tratamentos ativos A e B ou placebos A e B. Por exemplo, um investigador realizou um ensaio clínico fatorial para determinar se o uso a longo prazo de beta caroteno e ácido acetilsalicílico afetava o risco de câncer gastrintestinal.

Ensaio clínico randomizado cego. Delineamento no qual os participantes elegíveis são alocados aleatoriamente para os grupos de estudo com uma probabilidade predeterminada de ir para qualquer um deles e o grupo designado é ocultado aos investigadores, participantes e demais integrantes da equipe envolvida no estudo. Por exemplo, um ensaio clínico randomizado cego sobre um novo medicamento para o tratamento da diarreia exigiria que os participantes elegíveis fossem designados aleatoriamente ao novo medicamento ou a um comprimido idêntico placebo (geralmente com probabilidade de 50% de ser designado a um dos grupos) e que os investigadores, participantes e equipe do estudo não soubessem se o participante está tomando o medicamento ativo ou o placebo.

Ensaio clínico. Delineamento de pesquisa no qual os sujeitos recebem uma de (pelo menos) duas intervenções diferentes. Em geral, as intervenções são alocadas aleatoriamente, o que justifica o termo *ensaio clínico randomizado*. Os ensaios clínicos são às vezes denominados experimentos. Por exemplo,

um investigador realizou um ensaio clínico para avaliar se o tratamento profilático com penicilina reduzia o risco de endocardite bacteriana em pacientes com valvulopatias que eram submetidos a procedimentos dentários.

Epidemiologia. Ciência de determinar a frequência e os determinantes de doenças ou outros desfechos de saúde em populações. Por exemplo, um estudo investigou a epidemiologia da violência por armas de fogo em áreas de periferia.

Epidemiologista. Um médico, fortemente marcado pela idade e pelo sexo. Por exemplo, um dos autores deste livro (mas não será dito qual deles!)

Equipolência. Situação na qual não se sabe qual de duas possibilidades tem maior probabilidade de ser verdadeira (p. ex., o medicamento X é melhor do que o placebo, ou o medicamento X é pior do que o placebo). Portanto, é considerado ético comparar o medicamento X com o placebo em um ensaio clínico randomizado. Por exemplo, um grupo de investigadores acreditava que havia equipolência clínica em um ensaio clínico, pois não se sabia se um novo tratamento proposto para o câncer de esôfago iria resultar em desfechos melhores do que o cuidado-padrão atual.

Erro aleatório. Ocorre quando uma medida ou estimativa diverge do valor verdadeiro devido à variação ao acaso. O erro aleatório pode ser reduzido repetindo as aferições e aumentando o tamanho de amostra. Por exemplo, se a prevalência verdadeira do uso de óleo de peixe por pessoas com doença coronariana na população for de 20%, um estudo que arrola cem participantes poderia encontrar uma proporção de exatamente 20% de consumo de óleo de peixe, mas, simplesmente devido ao erro aleatório, é mais provável que essa proporção seja um pouco maior ou um pouco menor do que esse valor.

Erro de aferição (ou de medição). Situação na qual a precisão ou acurácia (ou ambas) de uma aferição são menos do que perfeitas; assim, a maioria das variáveis tem pelo menos um pouco de erro de aferição (com exceção, talvez, do óbito). Por exemplo, para reduzir o erro de aferição, um investigador utilizou um peso de 2 kg de aço inoxidável para calibrar a balança infantil semanalmente.

Erro de classificação. Erro de aferição para uma variável categórica no qual sujeitos com um valor da variável são contados (classificados erroneamente) como tendo outro valor. Por exemplo, investigadores estavam preocupados que devido à incompletude dos registros médicos, alguns sujeitos que sofreram uma queda durante sua hospitalização tinham sido classificados erroneamente como não tendo sofrido a queda. Ver também *erro de classificação diferencial* e erro de *classificação não diferencial*.

Erro sistemático. Ver *viés*.

Erro tipo I. Erro no qual uma hipótese nula que é verdadeira na população é rejeitada devido a um resultado estatisticamente significativo em um estudo. Por exemplo, um erro tipo I ocorre se um estudo sobre os efeitos do caroteno alimentar sobre o risco de desenvolver câncer de colo do intestino (com um alfa estipulado em 0,05) concluir que o caroteno reduz o risco de câncer de cólon ($P < 0,05$) quando na verdade não há associação. Ver também *resultado falso-positivo*.

Erro tipo II. Erro no qual uma hipótese nula que é falsa na população não é rejeitada por um estudo (isto é, $P >$ alfa). Por exemplo, ocorre um erro tipo II se um estudo deixa de rejeitar a hipótese nula de que o caroteno não tem efeito sobre o risco de câncer de colo do intestino (P > 0,05) quando na verdade o caroteno reduz o risco de câncer de colo do intestino. Ver também *resultado falso-negativo*.

Erro-padrão da média. Estimativa da precisão da média de uma variável contínua em uma amostra; depende do desvio padrão e (da raiz quadrada) do tamanho da amostra. Por exemplo, o investigador relatou que a idade média na coorte de 450 homens era de 59 anos, com um erro padrão de 0,48 anos.

Escala de Likert. Um conjunto de respostas (em geral, 5) a uma questão que fornece uma faixa de opções igualmente distribuídas. Por exemplo, as possíveis respostas para questão "Qual é a probabilidade de você retornar a este serviço de emergência para procurar atendimento?" foram as seguintes: Muito provável, Algo provável, Nem provável nem improvável, Algo improvável, Muito improvável.

Escala visual analógica (EVA). Escala (geralmente uma linha) que representa um espectro contínuo de respostas, de um extremo ao outro. Geralmente, a linha tem 10 cm de comprimento, e o escore é medido como a distância em centímetros a partir do extremo inferior. Por exemplo, uma escala visual analógica para a intensidade da dor pode apresentar uma linha reta com "sem dor" em um dos extremos e "dor insuportável" no outro; o participante do estudo marca um "X" no ponto que melhor descreve a intensidade de sua dor.

Escala. Abordagem comum para medir conceitos abstratos por meio de múltiplas questões utilizadas para atribuir um escore e combinadas em uma escala. Por exemplo, a escala SF36,[1] utilizada para medir qualidade de vida, inclui 36 questões que produzem 8 escalas relacionadas à saúde funcional e ao bem-estar. (SF vem do inglês "short form", que significa forma curta). Ver também *escala de Likert*.

Escore de propensão. Probabilidade estimada de que um participante do estudo terá um determinado valor de uma variável preditora, geralmente a probabilidade de receber um determinado tratamento. Controlar para que o escore de propensão (p.ex., por meio de pareamento, estratificação ou análise multivariável) seja um método para lidar com o confundimento por indicação. Ao invés de ajustar para todos os fatores que poderiam estar associados ao desfecho, o investigador cria um modelo multivariado para predizer quem irá receber o tratamento. Para cada sujeito é então atribuída uma estimativa da probabilidade de receber o tratamento (escore de propensão), que pode ser usada como a única variável confundidora ao estimar a associação entre o tratamento e o desfecho. Por exemplo, investigadores usaram um escore de propensão para ajustar para os fatores associados com o uso de aspirina, de modo a determinar a associação entre uso de ácido acetilsalicílico e câncer de cólon.

Especificação. Estratégia da fase de delineamento para lidar com um confundidor especificando um valor daquele confundidor como critério de inclusão para o estudo. Por exemplo, em um estudo sobre o efeito do uso de chupeta sobre o risco de síndrome da morte súbita do lactente, o investigador poderia usar a especificação para incluir no estudo apenas lactentes alimentados com fórmula. Se um menor risco de morte súbita for encontrado em usuários de chupeta, isso não poderia ser porque eles tinham maior probabilidade de serem amamentados no peito.

Especificidade. Proporção de sujeitos sem a doença que está sendo testada e nos quais o teste é negativo ("negativo na saúde" [NNS]). Por exemplo, comparando com resultados de biópsias, a especificidade de um PSA > 4,0 ng/mL é de aproximadamente 95% para a detecção de câncer de próstata; em outras palavras, 95% dos homens sem câncer de próstata terão um PSA ≤ 4,0 ng/mL. Ver também *sensibilidade*.

Estratificação. Estratégia de fase de análise para controlar o confundimento segregando os participantes do estudo em estratos de acordo com os níveis de um potencial confundidor e analisando a associação entre o preditor e o desfecho separadamente em cada estrato. Por exemplo, em um estudo sobre a associação entre exercício físico e risco de acidente vascular encefálico, não realizar atividade física regular poderia estar associado com um risco aumentado de acidente vascular encefálico porque muitas pessoas que não praticam exercício físico são obesas, e a obesidade aumenta o risco de acidente vascular encefálico. Para minimizar o potencial efeito confundidor da obesidade, os participantes foram estratificados de acordo com seu índice de massa corporal, e as análises foram realizadas separadamente naqueles que na linha de base tinham peso normal, sobrepeso ou obesidade.

Estudo tipo antes-depois. Um estudo que compara os atributos de sujeitos antes e após uma intervenção. Por exemplo, o estudo comparou a média dos níveis séricos de colesterol antes e depois de uma dieta com baixos teores de gordura.

Estudo analítico. Um estudo que busca associações entre duas ou mais variáveis. Por exemplo, o investigador realizou um estudo analítico sobre se a estatura estava correlacionada com a pressão arterial em estudantes de medicina. Ver também *estudo descritivo*.

Estudo cruzado de casos (*case-crossover study*). Variante do delineamento de caso-controle na qual cada caso serve como seu próprio controle, e o valor de uma exposição específica tempo-dependente no período anterior à ocorrência do desfecho é comparado com seu valor durante um ou mais períodos de tempo controle. Esse delineamento é suscetível ao viés recordatório e é, portanto, mais útil nas situações em que é possível avaliar uma exposição de forma objetiva. Por exemplo, um delineamento cruzado de casos foi usado para determinar se pacientes que consultavam em um serviço de emergência com crise de enxaqueca tinham maior probabilidade de terem comido chocolate nas duas horas anteriores do que em um período similar de tempo um dia antes.

Estudo cruzado. Delineamento no qual todos os sujeitos de um grupo de tratamento (ou controle) trocam para o outro grupo, geralmente na metade do estudo. Às vezes, há um período de *washout* (sem tratamento) entre as duas fases. Esse delineamento, que permite que todos os sujeitos recebam o tratamento ativo, é útil apenas para condições que voltam à situação da linha de base após o tratamento. Por exemplo, pacientes com enxaqueca participaram de um estudo cruzado que comparava um novo medicamento com placebo para a prevenção de enxaquecas.

Estudo de caso-controle aninhado. Estudo de caso-controle no qual os casos e os controles são selecionados a partir de uma coorte definida maior ou a partir de sujeitos previamente arrolados em um estudo de coorte. Esse delineamento é geralmente utilizado quando o custo de realizar certas aferições em todos os sujeitos da coorte é elevado demais; nesse delineamento, são feitas as aferições em material biológico armazenado na linha de base. Por exemplo, investigadores realizaram um estudo de caso-controle aninhado para avaliar se os níveis de citocinas em amostras de sangue de recém-nascidos estavam associados com o desenvolvimento de paralisia cerebral na coorte de nascimentos de 2009 do Estado de Ohio.

Estudo de caso-controle. Delineamento no qual os casos que têm uma doença (ou outro desfecho) são comparados com controles que não a têm. Por exemplo, um estudo de caso-controle comparou o consumo médio semanal de nozes e sementes em casos de diverticulite atendidos em um serviço de emergência com o consumo desses alimentos em controles que tinham outros diagnósticos de doenças digestivas.

Estudo de caso-coorte. Delineamento no qual os sujeitos que desenvolvem uma doença (ou outro desfecho) são selecionados como casos durante o seguimento de uma coorte maior e são então comparados com uma amostra aleatória de toda a coorte. Por exemplo, um estudo de caso-coorte arrolou uma coorte de 2 mil homens com câncer de próstata em estágio inicial e comparou os níveis de androgênios e vitamina D de amostras obtidas na linha de base naqueles que morreram durante o seguimento com os níveis em uma amostra aleatória de toda a coorte.

Estudo de coorte prospectiva. Delineamento no qual um grupo definido de participantes do estudo (a coorte) tem os valores basais de variáveis preditoras medidos e é então acompanhado ao longo do tempo para verificar o aparecimento de desfechos específicos. Por exemplo, o Nurses Health Study é um estudo de coorte prospectiva sobre fatores de risco para doenças comuns em mulheres. A coorte é uma amostra de enfermeiras nos Estados Unidos e os desfechos incluem doenças cardiovasculares, câncer e mortalidade.

Estudo de coorte retrospectiva. Estudo de coorte no qual a montagem da coorte, as medições basais e o seguimento ocorreram no passado. Por exemplo, para descrever a história natural de aneurismas da aorta torácica, um investigador que estava realizando um estudo de coorte retrospectiva em 2012, poderia obter dados de registros de altas hospitalares de pacientes com diagnóstico de aneurisma da aorta em 2007, e usar registros da alta hospitalar e de declarações de óbito para determinar quais pacientes tiveram ruptura do aneurisma aórtico ou morreram antes de 2012.

Estudo de coorte. Um estudo de coorte prospectiva envolve o arrolamento de um grupo de sujeitos (a coorte), a realização de algumas aferições na linha de base e então o seguimento dessa coorte ao longo do tempo para observar os desfechos. Um estudo de coorte retrospectiva envolve a identificação de um grupo de sujeitos (a coorte) nos quais as aferições já tenham sido feitas e todo ou parte do seguimento já tenha ocorrido. Por exemplo, um investigador realizou um estudo de coorte retrospectiva para avaliar se os resultados de um teste de inteligência emocional administrado quando soldados se alistaram no exército norte-americano estava associado com a probabilidade posterior de desenvolver transtorno do estresse pós-traumático (TEPT).

Estudo de coortes múltiplas. É um estudo de coorte que arrola dois ou mais grupos distintos de sujeitos (as coortes) e então compara seus desfechos. Frequentemente utilizado em estudos sobre exposições ocupacionais, nos quais as coortes que estão sendo comparadas foram ou não expostas a um potencial fator de risco. Por exemplo, os investigadores realizaram um estudo de coortes múltiplas sobre se a exposição a raios cósmicos durante voos de avião estava associada com um risco aumentado de doenças malignas hematológicas. Os investigadores estudaram quatro coortes: pilotos e comissários de bordo (que estariam expostos a raios cósmicos) e agentes de passagens e funcionários do portão de embarque (que não estariam). Ver também *estudo de dupla coorte*.

Estudo de dupla coorte. Delineamento no qual os sujeitos são arrolados em uma de duas coortes distintas, frequentemente definidas pela ocupação. Por exemplo, um estudo de dupla coorte foi usado para comparar os riscos de dermatite de contato nas mãos, assim como infecções fúngicas nos pés em ceramistas quando comparados com dançarinos.

Estudo de equivalência. Estudo cujo objetivo é mostrar que dois (ou mais) tratamentos têm desfechos similares. Em geral, um dos tratamentos é novo e o outro é sabidamente efetivo. Por exemplo, um estudo de equivalência foi realizado para comparar dois antibióticos (o medicamento novo A com o medicamento antigo B) para o tratamento da pneumonia.

Estudo descritivo. Estudo que não busca associações, não testa hipóteses ou realiza comparações. Por exemplo, o investigador realizou um estudo descritivo sobre a prevalência de obesidade em crianças pré-escolares. Ver também *estudo analítico*.

Estudo observacional. Termo genérico para um delineamento de pesquisa no qual os investigadores simplesmente observam os sujeitos sem realizar quaisquer intervenções. Assim, esse termo inclui estudos transversais, de caso-controle e de coorte, mas não ensaios clínicos randomizados ou estudos tipo antes-depois. Por exemplo, os examinadores realizaram um estudo observacional para determinar os fatores de risco para melanona.

Estudo pré-clínico. Estudo que ocorre antes de uma intervenção ser testada em humanos. Estes estudos podem incluir células, tecidos ou animais. Por exemplo, o US Food and Drug Administration exige estudos pré-clínicos em duas espécies de animais diferentes para documentar a segurança antes que novos tratamentos possam ser testados em humanos.

Estudo sobre reprodutibilidade. Estudo no qual a reprodutibilidade de uma medida é a questão de pesquisa principal. Ele é, em geral, realizado comparando os resultados de uma medida realizada múltiplas vezes pela mesma pessoa ou equipamento (reprodutibilidade intraobservador) ou os

resultados da mesma aferição feita por pessoas ou equipamentos diferentes (reprodutibilidade interobservador). Por exemplo, um grupo de investigadores realizou um estudo sobre reprodutibilidade para determinar se um novo estetoscópio eletrônico poderia melhorar a capacidade de detectar sopros diastólicos.

Estudo sobre teste diagnóstico. Estudo que avalia se os resultados de um procedimento médico são úteis para avaliar a probabilidade de um determinado diagnóstico em um paciente. Por exemplo, um estudo de teste diagnóstico foi delineado para determinar se os níveis séricos de bicarbonato eram úteis para diagnosticar sepse em pacientes com febre.

Estudo transversal. Delineamento no qual os sujeitos são selecionados e as aferições são feitas dentro de um período limitado de tempo, geralmente para estimar a prevalência de uma exposição ou de uma doença. Por exemplo, a prevalência de miopia foi estimada em um estudo transversal de 1.200 estudantes universitários em Berkeley, Califórnia.

Estudo-piloto. Pequeno estudo conduzido para determinar se um estudo maior é factível, assim como para otimizar a logística e maximizar a eficiência do estudo maior. Por exemplo, um ensaio clínico piloto sobre yoga restaurativo para a prevenção do diabetes em pacientes com resistência à insulina poderia buscar demonstrar a factibilidade de medir a resistência à insulina; refinar e padronizar a intervenção de yoga; e mostrar que é possível recrutar e randomizar participantes a grupos de yoga e controle.

Estudos sobre testes médicos. Termo genérico usado para estudos que medem quão bem um teste (ou uma série de testes) identifica pacientes com um determinado diagnóstico ou desfecho. Por exemplo, o investigador realizou um estudo sobre teste médico para determinar as razões de verossimilhança para a presença e ausência de angina típica (definida como dor ou pressão torácica subesternal aos esforços) no diagnóstico de doença arterial coronariana.

Experimento. Na pesquisa clínica, é um estudo no qual os sujeitos são alocados aleatoriamente para um (ou mais) grupo de tratamento ou comparação. Também é denominado ensaio clínico randomizado. Por exemplo, os investigadores realizaram um experimento para avaliar se o medicamento X era melhor do que o placebo para o tratamento da fibromialgia.

Exposição. Termo usado para indicar que um sujeito de um estudo tem um determinado fator de risco. Por exemplo, a exposição ao ácido acetilsalicílico foi definida como tomar uma média de um ou mais comprimidos de ácido acetilsalicílico (de qualquer dosagem) por semana durante o período anterior de seis meses.

Heterogeneidade. Situação na qual a associação entre uma variável preditora e uma de desfecho não é uniforme, seja entre diferentes estudos ou entre diferentes subgrupos de sujeitos. Por exemplo, há heterogeneidade substancial entre estudos que examinaram os efeitos do estrogênio na pós-menopausa sobre o humor e a cognição, com alguns estudos tendo mostrado efeitos positivos, alguns efeitos adversos e alguns não mostraram efeito.

Hiperpareamento. Situação na qual o pareamento além daquele necessário para controlar o confundimento reduz a capacidade do investigador de determinar se um fator de risco está associado com um desfecho, uma vez que os controles se tornaram excessivamente semelhantes aos casos. Por exemplo, uma vez que os controles foram pareados aos casos pela idade (± 3 anos), sexo, raça e situação socioeconômica, o hiperpareamento tornou impossível determinar se a escolaridade estava associada com o risco de acidente vascular encefálico em sujeitos com idade igual ou superior a 65 anos, pois as variáveis de pareamento são importantes determinantes da escolaridade naquele grupo etário.

Hipótese alternativa. É a proposição, utilizada no processo de estimar o tamanho de amostra, de que

existe uma associação na população entre a variável preditora e a de desfecho. Por exemplo, a hipótese alternativa do estudo era de que adolescentes que fumam têm uma probabilidade diferente de abandonar os estudos do que aqueles que não fumam.

Hipótese bilateral. Hipótese alternativa na qual o investigador está interessado em avaliar a possibilidade de incorrer em um erro tipo I nas duas direções possíveis (p.ex., risco maior ou risco menor). Por exemplo, um investigador testou a hipótese bilateral de que dançar salsa estava associado com um aumento ou uma diminuição no risco de demência. Ver também *hipótese unilateral*.

Hipótese complexa. Uma hipótese de pesquisa que tem mais de uma variável preditora ou de desfecho. Devem-se evitar hipóteses complexas, pois é difícil testá-las estatisticamente. Por exemplo, um grupo de investigadores reformulou sua hipótese complexa de que "um novo programa de gerenciamento de caso iria afetar tanto a duração da internação quanto a probabilidade de readmissão" em duas hipóteses simples ("Um novo programa de gerenciamento de caso iria afetar a duração da internação" e também que "Um novo programa de gerenciamento de caso iria afetar a probabilidade de readmissão"). Ver também *hipótese simples*.

Hipótese de pesquisa. Declaração formulada pelo investigador que sumariza os principais elementos do estudo, incluindo a população de interesse, as variáveis preditoras e de desfecho e um resultado antecipado. Para fins estatísticos, a hipótese de pesquisa é formulada de modo a estabelecer a base para os testes de significância estatística, geralmente incluindo uma hipótese nula e uma hipótese alternativa. Por exemplo, a hipótese de pesquisa era que a enxaqueca estaria associada com um aumento de pelo menos 20% no risco de acidente vascular encefálico.

Hipótese nula. É a forma da hipótese de pesquisa que especifica que não há diferença nos grupos que estão sendo comparados. Por exemplo, a hipótese nula declarava que o risco de desenvolver claudicação seria o mesmo em sujeitos com níveis normais de lipídeos e tratados com uma estatina do que naqueles tratados com placebo.

Hipótese simples. Hipótese com apenas uma variável preditora e uma variável de desfecho. Por exemplo, o investigador reformulou sua hipótese complexa em uma hipótese simples de que pessoas que comem frutas pelo menos cinco vezes por semana têm menor probabilidade de desenvolver câncer de colo do intestino. Ver também *hipótese complexa*.

Hipótese unilateral. Hipótese alternativa na qual o investigador está interessado em avaliar a possibilidade de incorrer em erro Tipo I em apenas uma das duas direções possíveis (p. ex., risco maior ou menor, mas não ambos). Por exemplo, um investigador testou a hipótese unilateral de que o tabagismo estava associado a um risco aumentado de demência. Ver também *hipótese bilateral*.

Hipótese. Termo genérico para uma declaração sobre o que se acredita que o estudo irá mostrar. Por exemplo, a hipótese de um estudo foi de que o uso crônico de antiepilépticos estava associado com um risco aumentado de câncer oral. Ver também *hipótese nula* e *hipótese de pesquisa*.

Hipóteses *post hoc*. Hipóteses que são formuladas após os dados já terem sido analisados. Por exemplo, em um estudo sobre a associação entre insônia e o risco de acidente vascular encefálico, a hipótese de que a insônia aumenta o risco de diverticulite é uma hipótese *post hoc*.

Homogeneidade. Situação na qual a associação entre uma variável preditora e uma de desfecho é uniforme em diferentes estudos. Por exemplo, há homogeneidade entre estudos de tamanho suficiente que examinaram o efeito do tabagismo sobre o câncer de pulmão. Todos encontraram um risco substancialmente aumentado em fumantes.

Identificador único. Coluna em uma tabela de um banco de dados relacional (melhor identificada por uma *chave principal*) que inclui informações sobre a pessoa, transação, resultado ou evento. Por

exemplo, uma tabela de Sujeitos poderia ter um registro para cada sujeito do estudo, sendo IdEstudo sua chave principal, assim como outras informações como campos com data de nascimento e sexo.

Incidência cumulativa. Ver *incidência*.

Incidência. Proporção de sujeitos que desenvolvem um desfecho durante o período de seguimento; às vezes denominada proporção de incidência ou incidência cumulativa. Por exemplo, um grupo de pesquisadores descobriu que gestantes vegetarianas tinham uma incidência menor de parto pré-termo do que mulheres que comiam carne.

Independente. Este termo é usado , pelo menos, de duas formas. Primeiramente, é a condição em que duas variáveis não sofrem influência uma da outra. Por exemplo, os investigadores determinaram que o consumo de nozes e a glicemia eram independentes: não havia evidência no seu estudo de que o consumo de nozes afetava a glicemia, ou vice versa. Em segundo lugar, independente se refere a um efeito que uma variável tem em outra variável e que não depende (isto é, "é independente de") uma terceira variável. Por exemplo, um investigador estava preocupado que a escolaridade materna estivesse associada ao aleitamento materno, e por isso ajustou para a escolaridade materna ao estimar o efeito independente do aleitamento materno sobre habilidades linguísticas de crianças com dois anos de idade.

Inferência. Processo de fazer conclusões sobre uma população com base em observações em uma amostra. Por exemplo, uma vez que o dobro do número de casos de câncer de bexiga apontava para o fato de os indivíduos relatarem tomarem água de poço quando comparados aos controles (P = 0,02), os investigadores inferiram que o consumo de água de poço aumenta o risco de câncer de bexiga na população.

Informações protegidas de saúde. Informações de saúde que permitem identificar o indivíduo. As normais federais norte-americanas sobre privacidade em saúde (denominadas HIPAA, Health Insurance Portability and Accountability Act) exigem que os pesquisadores mantenham a confidencialidade de informações protegidas de saúde ao realizarem suas pesquisas. Por exemplo, informações protegidas de saúde não devem ser armazenadas em *pendrives* ou enviadas por *e-mail* regular.

Inquérito. Estudo transversal em uma população específica, geralmente envolvendo um questionário. Por exemplo, o National Epidemiologic Survey on Alcohol and Related Conditions arrolou uma amostra representativa de adultos nos Estados Unidos e realizou perguntas sobre consumo de álcool no presente e no passado, transtornos relacionados ao uso de álcool e utilização de serviços para tratamento da dependência do álcool.

Interação. Outro nome para *modificação de efeito*.

Intervalo de confiança. Um termo frequentemente mal compreendido, o intervalo de confiança é uma medida de precisão: quanto mais estreito ele for, mais precisa será a estimativa. Os intervalos de confiança estão fortemente relacionados à significância estatística. Um intervalo de confiança de (1 – α)% inclui aproximadamente a faixa de valores que não foram significativamente diferentes em termos estatísticos (a um nível de significância de α) do que foi observado.

Intervenção. Em um ensaio clínico randomizado, é o tratamento ativo que os sujeitos recebem. Frequentemente utilizado como locução adjetiva (grupo de intervenção). Por exemplo, em um ensaio clínico randomizado sobre psicoterapia para o tratamento da ansiedade, a intervenção consistiu em seis meses de sessões semanais com duração de uma hora com psicólogo enfatizando o uso de técnicas cognitivo-comportamentais. Ver também *controle* (segunda definição).

Investigador principal. Pessoa que é responsável em última instância pelo delineamento e condução de um estudo e pela análise e apresentação dos achados. Por exemplo, o comitê de ética em pesquisa pediu para falar com o investigador principal do estudo porque alguns membros tinham dúvidas sobre o

protocolo.

Kapa. Termo estatístico que mede o grau em que dois (ou mais) observadores concordam ou não sobre a ocorrência de um fenômeno, além do que seria esperado pelo acaso. Varia de -1 (discordância total) a 1 (concordância total). Por exemplo, o kapa comparando o quanto dois patologistas concordavam sobre a presença de cirrose em uma amostra de lâminas de tecido hepático era de 0,85.

Má-conduta científica. Termo genérico para condutas que visam a enganar a comunidade científica, incluindo má conduta em pesquisa (fabricação, falsificação de dados e plágio), assim como autoria honorária e fantasma e conflitos de interesses não relatados ou manejados. Por exemplo, a instituição do investigador considerou que ele era culpado por má conduta científica porque não relatou que tinha ações da empresa que fabricou o equipamento médico que ele estava estudando.

Má-conduta em pesquisa. Conduta ilegal ou antiética em um estudo, incluindo plágio e fabricação ou falsificação de dados de pesquisa. Por exemplo, descobriu-se que um coordenador de pesquisa no VA Medical Center, em Albany, Nova Iorque, repetidamente submetia documentações falsas para permitir que pessoas que não se qualificavam para um estudo fossem arroladas. Todos os dados do centro de Albany foram posteriormente excluídos, de modo que o tempo e os esforços dos participantes foram desperdiçados. Ver também *má conduta científica*.

Magnitude de efeito. No contexto do planejamento do tamanho de amostra, é uma medida do tamanho da diferença que o investigador quer detectar entre os grupos que serão comparados, ou do tamanho da associação. Por exemplo, um grupo de investigadores baseou suas estimativas de tamanho de amostra em uma magnitude de efeito correspondente a uma diferença na glicemia entre dois grupos de 20 mg/dL.

Marcador substituto. Medida que se imagina estar associada com desfechos clinicamente relevantes. Um bom marcador substituto geralmente mede mudanças em um fator intermediário na rota principal que determina o desfecho clínico. Por exemplo, um aumento na contagem de linfócitos CD4 em pacientes com infecção pelo vírus da imunodeficiência humana (HIV) é um bom marcador substituto para a efetividade de antirretrovirais, pois prediz um risco mais baixo de infecções oportunistas.

Mascaramento. Ver *cegamento*.

Média. O valor médio de uma variável contínua em uma amostra ou população; calculado como a soma de todos os valores daquela variável dividida pelo número de sujeitos. Por exemplo, a média dos níveis séricos de colesterol em uma amostra de 287 mulheres de meia idade era de 223 mg/dL. Ver também *mediana* e *desvio-padrão*.

Mediador. Uma variável que é causada pelo preditor de interesse e também causa o desfecho; ela responde, pelo menos em parte, por como o preditor causa o desfecho. Por exemplo, ao estudar o efeito da obesidade sobre o risco de acidente vascular encefálico, os investigadores não controlaram para o diabetes, pois eles acreditavam que um mecanismo pelo qual a obesidade poderia levar ao acidente vascular encefálico era como mediador causando o diabetes.

Mediana. Valor de uma variável que divide uma amostra ou população em duas metades de tamanho (aproximadamente) igual. Equivalente ao percentil 50. Com frequência, utilizada quando uma variável contínua tem alguns poucos valores muito elevados (ou muito baixos) que podem influenciar excessivamente a média. Por exemplo, a mediana da renda anual na amostra de 54 médicos era de $225.000. Ver também *média* e *desvio-padrão*.

Medições pareadas. Medições fortemente relacionadas umas com as outras de alguma forma, como aquelas realizadas em lados diferentes da mesma pessoa, diferentes membros de um par de gêmeos ou (o que é mais comum) o mesmo participante em dois momentos diferentes, como antes e depois de uma

intervenção. Por exemplo, em um estudo sobre o efeito de um programa de exercícios físicos sobre os níveis de hemoglobina glicosilada em pacientes com diabetes tipo II, medições pareadas da hemoglobina glicosilada incluíram medidas feitas na linha de base e depois de três meses de exercício físico.

Metanálise. Processo de combinar os resultados de diversos estudos com variáveis preditoras e de desfecho semelhantes em um único resultado-sumário. Por exemplo, uma metanálise de 12 estudos publicados mostrou que o uso de anti-inflamatórios não esteroides esteve associado com um risco 28% maior de desenvolver asma.

Modelo de Cox. Também denominado *modelo de azares proporcionais de Cox*. Técnica estatística multivariada que mede os efeitos individuais de uma ou mais variáveis preditoras sobre a taxa (azar) em que um desfecho ocorre em uma amostra, levando em consideração as durações diferentes do seguimento entre os sujeitos. Por exemplo, utilizando um modelo de azares proporcionais de Cox, homens tiveram o dobro da probabilidade de mulheres, e negros, o triplo da probabilidade de brancos de desenvolver acidentes vasculares encefálicos, após ajuste para idade, pressão arterial, diabetes e duração do seguimento. Ver também *modelo de regressão logística*.

Modelo de efeitos aleatórios. Termo genérico utilizado na análise estatística multinível; é utilizado neste livro apenas no que se refere à metanálise, em que descreve um modelo estatístico no qual os pesos dos estudos e a variância da estimativa-sumário de efeito incorporam um termo para a variabilidade entre os resultados dos estudos individuais que foram incluídos. Por exemplo, em uma metanálise de ensaios clínicos sobre o efeito da prática de ioga sobre a depressão, os resultados dos ensaios clínicos foram variados; assim, os estudos menores contribuíram mais para o efeito sumário baseado no modelo de efeitos aleatórios, e o intervalo de confiança foi mais amplo do que quando foi utilizado o modelo de efeitos fixos. Ver também *modelo de efeitos fixos*.

Modelo de efeitos fixos. Termo genérico usado na análise estatística multinível; discutido neste livro apenas no contexto da metanálise, onde descreve um modelo estatístico no qual os pesos dos estudos e a variância da estimativa sumária de efeito se baseiam apenas nas variâncias intraestudos dos estudos incluídos. Por exemplo, em uma metanálise de ensaios clínicos sobre o efeito de praticar ioga na depressão, os resultados dos ensaios clínicos foram variados; o efeito-sumário com base no modelo de efeitos fixos foi dominado por um estudo de grande porte, e o intervalo de confiança foi mais estreito do que teria sido se fosse utilizado o modelo de efeitos aleatórios. Ver também *modelo de efeitos aleatórios*.

Modelo de regressão logística. Técnica estatística usada para estimar os efeitos de uma ou mais variáveis preditoras sobre uma variável de desfecho dicotômica, ajustando para os efeitos de outras variáveis preditoras e confundidoras. Por exemplo, em um modelo de regressão logística, os homens tiveram o dobro da probabilidade das mulheres, e pessoas negras, o triplo da probabilidade de pessoas brancas, de desenvolver acidente vascular encefálico, após ajuste para idade, pressão arterial e diabetes.

Modificação de efeito. Condição na qual a força da associação entre uma variável preditora e uma variável de desfecho é afetada por uma terceira variável (frequentemente denominada modificadora de efeito, embora possa ser difícil determinar qual é a variável preditora e qual é a modificadora de efeito). Por exemplo, investigadores mostraram que os efeitos da renda sobre o risco de acidente vascular encefálico eram diferentes em pessoas brancas e negras; portanto, a pobreza tinha uma associação mais forte com acidente vascular em negros do que em brancos. Ver também *confundimento*.

Nível de significância estatística. Ver *alfa*.

Normalização. Em um banco de dados relacional, é o processo de eliminar a redundância e melhorar a legibilidade assegurando-se de que cada dado seja armazenado apenas nas linhas ou tabelas necessárias.

Por exemplo, após o banco de dados ter sido normalizado por um consultor de banco de dados, ele conseguiu atualizar o número de telefone de um sujeito apenas alterando uma única linha em uma única tabela.

Número necessário tratar. Número absoluto de pessoas que recebem um tratamento para prevenir a ocorrência de um desfecho. Calculado como a recíproca da diferença de riscos. Por exemplo, ao avaliar os benefícios de tratar a hipertensão leve a moderada, o número necessário tratar era de 800 pacientes por ano para prevenir um acidente vascular encefálico.

Objetivos específicos. Em uma proposta de pesquisa, declarações breves sobre os objetivos da pesquisa. Por exemplo, um objetivo específico de um ensaio clínico sobre o efeito da testosterona na densidade mineral óssea em homens poderia ser: "Testar a hipótese de que, comparados com homens designados para receber adesivo placebo, aqueles designados para receber o adesivo de testosterona terão menor perda óssea durante três anos de tratamento".

Padrão-ouro. Método não ambíguo de determinar se um paciente tem ou não uma determinada doença ou desfecho. Por exemplo, o padrão-ouro para o diagnóstico de fratura de quadril exigia confirmação radiológica por um radiologista certificado.

Padronização. Instruções específicas e detalhadas sobre como realizar uma aferição, com o objetivo de maximizar a reprodutibilidade e a precisão da aferição. Por exemplo, em um estudo que mede a pressão arterial, a padronização dessa aferição poderia incluir instruções sobre como preparar o participante, que tamanho de manguito usar, onde posicionar o manguito, o quanto insuflar e desinsuflar o manguito e quais ruídos cardíacos indicam a pressão sistólica e a diastólica.

Pareamento. Em um estudo de caso-controle, é o processo de selecionar controles que sejam similares em relação a certos atributos aos casos, para reduzir o confundimento por esses atributos. Por exemplo, em um estudo de caso-controle sobre os fatores de risco para brucelose, os controles foram pareados aos casos por idade (intervalo de três anos), sexo e município de residência. Ver também *hiperpareamento*.

Particionamento recursivo. Técnica multivariada para classificar as pessoas de acordo com seu risco de um desfecho; ao contrário de técnicas que exigem um modelo, como a regressão logística, o particionamento recursivo não requer pressupostos sobre a forma da relação entre as variáveis preditoras e de desfecho. Ele cria uma árvore de classificação que inclui uma série que questões do tipo sim/não, denominada Árvore de Classificação e Regressão (CART, *Classification and Regression Tree*). Por exemplo, por meio do particionamento recursivo, um grupo de pesquisadores determinou que pacientes de 20 a 65 anos de idade, que consultavam em serviço de emergência por dor abdominal mas não tinham perda de apetite, febre ou dor à descompressão súbita, tinham baixo risco de apendicite. Ver *regra de predição clínica* e *sobreajuste*.

Participante. Alguém que participa em um estudo. O termo participante muitas vezes é preferido em relação a sujeito, pois enfatiza que a pessoa arrolada no estudo é um participante ativo para o avanço da ciência, e não meramente um sujeito que está sendo examinado. Por exemplo, em um estudo sobre um novo medicamento para o tratamento da insônia, os participantes são as pessoas elegíveis para um estudo e que são recrutadas para ele.

Participantes vulneráveis. Potenciais participantes do estudo que estão em maior risco para serem usados de forma eticamente inadequada na pesquisa. Por exemplo, uma vez que pessoas com limitação cognitiva ou problemas de comunicação podem ser incapazes de fornecer o consentimento informado completo para participar na pesquisa, eles são considerados pessoas vulneráveis. Outros exemplos incluem crianças, presidiários, fetos e pessoas em situação socioeconômica desfavorável.

Período de *run-in*. Em um ensaio clínico, é o breve período durante o qual todos os participantes

elegíveis recebem placebo ou a intervenção ativa; apenas aqueles que alcançaram um determinado nível de adesão, toleraram a intervenção ou tiveram um desfecho intermediário são elegíveis para o ensaio clínico principal. Por exemplo, no *Cardiac Arrhythmia Suppression Trial*, apenas os participantes que tiveram redução satisfatória nas extrassístoles ventriculares enquanto usavam o medicamento ativo durante o período *run-in* foram randomizados para continuarem usando o medicamento ou trocarem para placebo.

Período de *washout*. Em um estudo cruzado, é o período de tempo entre o primeiro e o segundo tratamento, para permitir que os efeitos da intervenção desapareçam e que a medida do desfecho retorne aos níveis basais. Por exemplo, em um ensaio clínico cruzado comparando um diurético contra um placebo para o tratamento da hipertensão arterial sistêmica, o investigador poderia permitir um período de *washout* de um mês sem tratamento entre os dois períodos de tratamento, para permitir que a pressão arterial retorne aos níveis basais.

Pesquisa translacional. Pesquisa que busca traduzir achados científicos de modo a melhorar a saúde. A pesquisa translacional pode ter como objetivo testar achados da ciência básica produzidos em laboratório em estudos clínicos com pacientes (muitas vezes chamado de “pesquisa da bancada ao leito” ou “pesquisa T1”) ou aplicar os achados de estudos clínicos para melhorar a saúde de populações (muitas vezes chamado de “pesquisa do leito à população” ou “pesquisa T2”). Por exemplo, um estudo para determinar se um defeito genético que causa surdez congênita em ratos tem efeito semelhante em humanos seria um estudo de pesquisa T1; um estudo para determinar se esforços em todo o estado para rastrear recém-nascidos com um teste que mede a resposta cortical ao som para detectar a perda auditiva melhora o desempenho escolar seria um estudo de pesquisa T2.

Pessoa-tempo. Soma das durações de tempo em que cada um dos sujeitos em um estudo ou população está em risco, usada como denominador para calcular as taxas de incidência. Pode ser calculada como o número de sujeitos que estão em risco de um desfecho multiplicado pelo seu tempo médio em risco. Por exemplo, o total de pessoa-tempo de seguimento entre os 1000 sujeitos que tiveram um período médio em risco de 2,5 anos foi um total de 2500 pessoas-ano, embora 5% dos sujeitos tenham sido seguidos por um mês ou menos. Ver também *taxa de incidência*.

Placebo. Ver *controle placebo*.

Plágio. Tipo de má conduta científica no qual um investigador se apropria de ideias, resultados ou palavras de uma outra pessoa sem dar o devido crédito. Por exemplo, utilizar a descrição de outro investigador sobre um novo método de aferição sem dar o devido crédito constitui plágio.

Poder estatístico. Probabilidade de rejeitar corretamente a hipótese nula em uma amostra se o efeito real na população for igual ou superior a uma magnitude de efeito especificada. Por exemplo, suponha que o exercício físico leve a uma redução média de 20 mg/dL nos níveis de glicemia de jejum em mulheres diabéticas na população geral. Se um investigador fixou o poder estatístico em 90% e selecionou uma amostra da população em diversas ocasiões, sempre realizando o mesmo estudo com as mesmas aferições, então, em 9 de cada 10 estudos, o investigador iria rejeitar corretamente a hipótese nula e concluir que o exercício físico reduz os níveis glicêmicos. Ver também *beta*.

População acessível. Grupo de pessoas às quais o investigador tem acesso e que poderão ser selecionadas ou convidadas para participar no estudo. Por exemplo, a população acessível para um estudo consistia em mulheres com câncer de mama cujo tratamento iniciou até seis semanas após o diagnóstico, com acompanhamento no Longview Hospital entre 1º de janeiro de 2013 e 30 de junho de 2014. Ver também *amostra pretendida* e *população alvo*.

População. Conjunto completo de pessoas com características específicas. Por exemplo, a população adulta nos Estados Unidos com diabetes tipo 2 poderia ser definida como todos os adultos dos estados

únicos que tomam antidiabético oral ou que têm uma glicemia acima de 125 mg/dL.

População-alvo. Grande conjunto de pessoas definidas por características clínicas e demográficas, para o qual o investigador deseja generalizar os achados de um estudo. Por exemplo, a população-alvo para o estudo de um novo tratamento para asma em crianças no hospital do investigador poderia ser as crianças com asma em todo o mundo.

Precisão. Grau em que uma medida de uma variável é reprodutível, com quase o mesmo valor cada vez que é aferida. Por exemplo, uma balança pode medir o peso corporal com grande precisão, enquanto uma entrevista para medir a gravidade da depressão provavelmente irá produzir valores que variam dependendo do observador.

Pré-teste. Avaliação de questionários, medidas ou procedimentos específicos que pode ser realizada pela equipe do estudo antes dele iniciar. O objetivo é avaliar a funcionalidade, a adequação ou a factibilidade dessas medidas. Por exemplo, para pré-testar o sistema de entrada de dados e de gerenciamento do banco de dados, pode-se solicitar que a equipe do estudo preencha formulários deixando dados faltantes, fora da faixa permitida ou ilógicos, para verificar se o sistema de edição de dados é capaz de identificar esses erros.

Prevalência. Proporção de pessoas com uma doença ou condição em um determinado ponto no tempo. Por exemplo, a prevalência de lúpus eritematoso sistêmico é a proporção de pessoas que têm essa doença em um ponto específico no tempo; ela pode aumentar se a doença se tornar mais comum ou se o tratamento melhorar de forma que pessoas com a doença vivam mais tempo.

Proposta de pesquisa. Documento escrito para fins de obtenção de financiamento de pesquisa e que descreve o delineamento proposto, os participantes, as aferições, as análises estatísticas e as questões éticas. Por exemplo, o National Institutes of Health recebe milhares de propostas de pesquisa anualmente de investigadores que buscam financiamento para seus estudos.

Proposta. Documento que inclui o protocolo de um estudo, o orçamento e outras informações administrativas e de apoio que são redigidas com o objetivo de obter financiamento de uma agência de fomento à pesquisa. Por exemplo, o National Institutes of Health (NIH) exige que propostas de financiamento para muitos tipos de pesquisa.

Protocolo. Plano detalhado por escrito de um estudo. Por exemplo, o protocolo de um estudo especificou que apenas os sujeitos que compreendiam inglês no nível esperado para a oitava série eram elegíveis para participação.

Questão de pesquisa secundária. Questão que não seja a questão de pesquisa principal, muitas vezes incluindo preditores ou desfechos adicionais. Por exemplo, se a questão de pesquisa principal for determinar a associação do consumo de álcool em gestantes com a ocorrência de baixo peso ao nascer, uma questão secundária poderia ser determinar a associação do consumo de álcool com a ocorrência de anemia durante a gestação.

Questão de pesquisa. Questão que um projeto de pesquisa busca responder. Uma boa questão de pesquisa deveria incluir o preditor e o desfecho de interesse, além da população que será estudada. Questões de pesquisa geralmente são formuladas como “A está associado com B na população C?” ou (para um ensaio clínico) “A causa B na população C?”. Por exemplo, “O uso regular de fio dental reduz o risco de eventos coronarianos em pessoas com diabetes?”.

Questionário. Instrumento de aferição que consiste em uma série de perguntas para obter informações dos participantes do estudo. Os questionários podem ser autoadministrados ou administrados pela equipe do estudo. Por exemplo, o Block Food Frequency Questionnaire pergunta sobre a ingesta usual de 110 itens alimentares para avaliar a ingestão de múltiplos nutrientes e grupos alimentares.

Randomização em blocos e estratificada. Procedimento de randomização com o objetivo de assegurar que números iguais de participantes com uma determinada característica (geralmente um confundidor) sejam alocados aleatoriamente para cada um dos grupos de estudo. A randomização é estratificada de acordo com a característica de interesse; dentro de cada estrato, os participantes são designados aleatoriamente em blocos de tamanho predeterminado. Por exemplo, em um ensaio clínico sobre um medicamento para prevenir fraturas, uma história de fratura vertebral é um preditor tão forte do desfecho e da resposta a muitos tratamentos que é melhor assegurar um número igual de participantes com e sem fratura vertebral em cada um dos grupos do estudo. Portanto, os investigadores usaram a randomização em blocos e estratificada para dividir os participantes em dois estratos (aqueles com fraturas vertebrais e aqueles sem essas fraturas); dentro de cada estrato, a randomização foi realizada em blocos de seis a dez sujeitos.

Randomização em blocos. Um método de alocar sujeitos a uma determinada intervenção em blocos (grupos) de tamanho pré-especificado (p.ex., 4 ou 6), de modo a assegurar que números similares de sujeitos sejam alocados aos grupos de intervenção e controle. Frequentemente utilizado em estudos multicêntricos nos quais os investigadores querem que os números totais de sujeitos randomizados para intervenção ou controle sejam semelhantes em cada centro. Por exemplo, os pacientes em cada clínica foram alocados aleatoriamente aos grupos de tratamento ou controle em blocos de seis, garantindo que o número de sujeitos por grupo iria diferir por não mais do que três. Ver também *randomização em blocos e estratificada*.

Randomização mendeliana. Técnica para fortalecer a inferência causal aproveitando-se da herança aleatória de genes que afetam a suscetibilidade a um fator de risco ou a um tratamento. Por exemplo, a probabilidade de uma relação causal entre uso materno de paracetamol e asma em crianças foi fortalecida pela observação de que a associação era significativamente mais forte em mãos com o genótico T1 da glutationa S-transferase, uma enzima envolvida na detoxificação de um metabólito do paracetamol.

Randomização por conglomerados. Técnica em que grupos de participantes, denominados conglomerados, são alocados aleatoriamente a diferentes tratamentos, em vez de cada participante ser alocado aleatoriamente como um indivíduo. Por exemplo, em um estudo sobre os efeitos da redução de ruídos sobre a recuperação após uma cirurgia cardíaca, o investigador utilizou a randomização por conglomerados para alocar unidades de terapia intensiva em 40 hospitais diferentes para uma "intervenção pós-operatória silenciosa" ou para um grupo-controle de "cuidado usual".

Randomização. Processo de alocar aleatoriamente participantes elegíveis a um dos grupos de estudo em um ensaio clínico randomizado. O número de grupos de tratamento e a probabilidade de ser designado a um dos grupos são determinados antes de iniciar a randomização. Embora os participantes elegíveis sejam geralmente designados a dois grupos de estudo com probabilidade igual (50%), a alocação aleatória pode ser feita a qualquer número de grupos de estudo com qualquer probabilidade predeterminada. Por exemplo, em um estudo que comparava dois tratamentos com um controle placebo, a randomização poderia ocorrer para três grupos, com uma probabilidade predeterminada de 30% de ser para qualquer um dos dois grupos de tratamento ativo e de 40% de ser para o grupo placebo.

Razão de azares. A razão entre a taxa de azares de pessoas expostas a um fator de risco e a taxa de azares em pessoas não exposta. É quase sempre estimada a partir de um modelo de azares proporcionais (modelo de Cox). Por exemplo, a razão de azares para desenvolver doença arterial coronariana era de 2,0 quando se comparavam homens de 50 a 59 anos com mulheres na mesma faixa etária.

Razão de chances. É a razão das chances de uma doença (ou de outro desfecho) naqueles expostos a um fator de risco contra as chances da mesma doença naqueles não expostos. A razão de riscos e a

razão de chances são semelhantes quando uma doença é rara tanto nos indivíduos expostos quanto naqueles não expostos, pois as chances e os riscos da doença são similares. Por exemplo, a razão de chances para a insuficiência renal entre pessoas com hipertensão é de 2,0, o que significa que pessoas hipertensas têm uma chance duas vezes maior de desenvolver insuficiência renal do que pessoas não hipertensas.

Razão de riscos (risco relativo). Risco de um desfecho em um grupo dividido pelo risco em um grupo de comparação. Por exemplo, se o risco de eventos tromboembólicos em mulheres usuárias atuais de estrogênio for de 5/1000 (0,5%) e o risco naquelas que nunca usaram estrogênio for de 2/1000 (0,2%), o risco relativo em mulheres usando estrogênio quando comparadas às não usuárias é de 2,5. Ver também *razão de azares* e *razão de chances*.

Razão de verossimilhança. Termo usado para descrever os efeitos quantitativos de um resultado de teste médico sobre a probabilidade de que um paciente tenha a doença para a qual está sendo testado. É definida como a probabilidade (verossimilhança) daquele resultado de teste em um paciente com a doença dividida pela probabilidade (verossimilhança) do mesmo resultado em um paciente sem a doença (no inglês, é lembrada pelo mnemônico WOWO: *with over without*, ou seja, com sobre sem). Por exemplo, a razão de verossimilhança para os sintomas característicos de angina típica (pressão subesternal aos esforços) é em torno de 50 para o diagnóstico de doença arterial coronariana.

Recrutamento. Processo de identificar e arrolar participantes elegíveis em um estudo. Os métodos de recrutamento variam dependendo da natureza do estudo. Por exemplo, o recrutamento para um estudo incluía identificar sujeitos elegíveis em ambulatórios especializados, anúncios em filipetas e jornais e o uso da internet e de mídias sociais.

Redes de pesquisa baseadas em serviços clínicos. Redes nas quais médicos que trabalham na comunidade se juntam para estudar questões de pesquisa de interesse. Por exemplo, um estudo de uma rede de pesquisa baseada em serviços clínicos sobre tratamentos para síndrome do túnel do carpo na atenção primária à saude mostrou que a maioria dos pacientes melhoravam com o tratamento conservador. Isso contrastava com a literatura prévia de centros médicos acadêmicos que mostrava que a maioria dos pacientes com síndrome do túnel do carpo precisava de cirurgia.

Registro. Banco de dados de pessoas com uma determinada doença ou que foram submetidas a um determinado procedimento. Estudos podem ser conduzidos utilizando registros por meio da coleta de dados dos desfechos como parte do registro ou relacionando dados do registro com outras fontes, como registros de câncer ou registros de óbitos. Por exemplo, o San Francisco Mammography Registry obtém dados de todas as mulheres submetidas a mamografias nos três maiores centros de mamografia em São Francisco; investigadores relacionaram os dados desse registro com registros locais de câncer para estimar a acurácia da mamografia.

Regra de predição clínica. Algoritmo que combina diversos preditores, incluindo a presença ou a ausência de vários sinais e sintomas e os resultados de testes médicos, para estimar a probabilidade de ocorrência de uma determinada doença ou desfecho. Por exemplo, investigadores desenvolveram uma regra de predição clínica para o diagnóstico de fraturas do punho em mulheres pós-menopáusicas com base em informações sobre fraturas prévias, características da queda (se tiver ocorrido), exame físico do antebraço e medicamentos em uso.

Regressão à média. Tendência de valores extremos (muito elevados ou muito baixos) se aproximarem da média populacional quando medidos novamente. Por exemplo, em um grupo de crianças selecionadas para um estudo por terem pressão sistólica acima do percentil 95, a maioria delas apresentou valores pressóricos mais baixos na primeira visita de seguimento, mesmo não tendo recebido nenhum tratamento.

Resultado falso-negativo. Termo que pode ser usado de duas formas diferentes. No contexto de um teste médico, refere-se a um resultado de teste que é falsamente negativo em um paciente com a condição que está sendo testada. Por exemplo, embora a paciente tivesse câncer de mama comprovado por biópsia, sua mamografia havia apresentado um resultado falso-negativo. No contexto de um estudo de pesquisa, refere-se a um resultado de estudo que deixa de detectar um efeito na amostra (p. ex., o resultado do estudo não foi estatisticamente significativo) que está presente na população. Por exemplo, embora estudos subsequentes tenham mostrado que o tabagismo aumente o risco de acidente vascular encefálico, um estudo de caso-controle inicial tinha apresentado um resultado falso-negativo (P = 0,23).

Resultado falso-positivo. Termo que pode ser usado de duas formas diferentes. No contexto de um teste médico, refere-se a um resultado de teste que é falsamente positivo em um paciente que não tem a condição que está sendo testada. Por exemplo, embora a paciente não tivesse tido câncer de mama nem desenvolvido essa doença durante o seguimento de seis anos, sua mamografia havia apresentado um resultado falso-positivo. No contexto de um estudo de pesquisa, refere-se a um resultado de um estudo que detecta um efeito na amostra (p. ex., resultado estatisticamente significativo) que não está presente na população. Por exemplo, embora estudos subsequentes tenham mostrado que o tabagismo não aumenta o risco de doença de Parkinson, um estudo de caso-controle havia inicialmente apresentado um resultado falso-positivo (P = 0,03).

Revisão por pares. Revisão de um protocolo, proposta ou manuscrito por pares do investigador que preparou esses documentos. Por exemplo, propostas submetidas para financiamento pelo NIH passam por um processo de revisão por pares em que cientistas da mesma área atribuem escores ao protocolo utilizando critérios bem definidos. Da mesma forma, manuscritos submetidos a revistas médicas passam por revisão por pares por cientistas que ajudam editores a decidir se o manuscrito deve ser publicado.

Revisão sistemática. Revisão da literatura médica que utiliza uma abordagem sistemática para identificar todos os estudos sobre uma determinada questão de pesquisa, com critérios claros para incluir um estudo na revisão e métodos padronizados para extrair os dados dos estudos incluídos. Uma revisão sistemática pode também incluir uma metanálise dos resultados dos estudos. Por exemplo, o investigador realizou uma revisão sistemática de todos os estudos que testaram se suplementos de zinco reduziam o risco de desenvolver resfriados.

Risco relativo. Ver *razão de riscos*.

Sensibilidade. Proporção de sujeitos com a doença nos quais um teste é positivo (“positivo na doença” [PND]). Por exemplo, comparado com os resultados da biópsia, a sensibilidade do PSA sérico > 4,0 ng/mL é de aproximadamente 20% para a detecção de câncer de próstata; em outras palavras, 20% dos homens com câncer de próstata terão um PSA > 4,0 ng/mL. Ver também *especificidade*.

Sobreajuste. Problema que surge quando os investigadores selecionam variáveis ou pontos de corte para um modelo multivariado com base, em parte, na variação ao acaso na amostra, levando a uma baixa capacidade de generalização dos resultados. Por exemplo, revisores suspeitaram de sobreajuste quando os autores relataram que o melhor modelo para predizer cataratas recorrentes incluía ter nascido nos meses de março ou agosto para mulheres entre 65 e 74 anos.

Sujeito. Ver *participante*.

Supressão. Tipo de confundimento no qual a variável confundidora diminui a associação aparente entre a variável preditora e a variável de desfecho porque está associada com a variável preditora, porém afeta o desfecho na direção oposta. Por exemplo, uma associação entre tabagismo e rugas na pele poderia passar despercebido em um estudo (ser “suprimida”) se os fumantes forem mais jovens e o confundimento pela idade não for controlado.

Tabela de dados. Tabela dos dados do estudo na qual cada linha corresponde a um único registro e cada coluna corresponde a um campo ou atributo. Todos os estudos têm uma tabela de sujeitos na qual cada linha corresponde a um participante individual e as colunas correspondem a informações específicas de cada participante, como sexo e data de nascimento. A maioria dos estudos também utilizam tabelas adicionais, nas quais as linhas correspondem a visitas do estudo, a resultados laboratoriais, a contatos telefônicos, etc.

Tamanho de amostra. Esse termo tem dois significados. Pode ser o número de participantes arrolados em um estudo ou o número estimado de participantes necessários para um estudo ser bem sucedido. Por exemplo, o investigador estimou que precisaria de um tamanho de amostra de 54 sujeitos para ter um poder estatístico de 90% para detectar um risco duas vezes maior de comportamento agressivo em meninos da terceira série expostos a jogos violentos de videogame.

Taxa de azares. Termo epidemiológico que mede a taxa instantânea em que um desfecho ocorre em uma população. Para fins práticos, é quase sempre estimada como a taxa de um desfecho. Por exemplo, a taxa de azares para desenvolver doença arterial coronariana em mulheres entre 50 a 59 anos foi estimada em 0,008 por ano.

Taxa de incidência. Taxa em que uma determinada doença ou desfecho ocorrem em um grupo de sujeitos anteriormente livres daquela doença. Em geral, á calculada como o número de novos casos do desfecho dividido pelo número de pessoas-tempo em risco. Por exemplo, a taxa de incidência de infarto do miocárdio era de 35,3 por 1.000 pessoas-ano em homens de meia idade, em torno de duas vezes a taxa em mulheres de meia idade (17,4 por mil pessoas-ano). Ver também *pessoa-tempo*.

Taxa de resposta. Proporção de participantes elegíveis que respondem a um questionário ou a um determinado item dele. Uma baixa taxa de resposta pode reduzir a validade interna do estudo e enviesar o desfecho. Por exemplo, em um inquérito com estudantes do ensino médio, uma taxa de resposta de 20% a uma questão sobre consumo de maconha sugere que o resultado provavelmente não é uma estimativa válida da taxa real de uso de maconha entre estudantes. Ver também *dados faltantes*.

Taxa. Uma medida do risco, definida como o número de sujeitos que desenvolvem um desfecho dividido pelas pessoas-tempo em risco. Por exemplo, a taxa de acidente vascular encefálico no estudo foi 23 por 1.000 pessoas-ano. Ver também *taxa de azares*.

Teste de hipóteses múltiplas. Situação na qual um investigador estuda mais de uma – e geralmente muito mais de uma – hipótese em um estudo, aumentando, assim, o risco de incorrer em um erro Tipo I, a não ser que o nível de significância estatística seja ajustado. Por exemplo, embora o investigador tenha relatado uma associação estatisticamente significativa ($P = 0{,}03$) entre consumo de vitamina D e declínio cognitivo, os seus resultados foram criticados porque ele não levou em conta o efeito do teste de hipóteses múltiplas, pois o estudo havia examinado mais de 30 suplementos nutricionais. Ver também *correção de Bonferroni*.

Teste do qui-quadrado. Técnica estatística que compara duas (ou mais) proporções para determinar se elas são estatisticamente diferentes uma da outra. Por exemplo, um estudo determinou se o risco de demência era semelhante em pessoas que faziam exercícios físicos pelo menos duas vezes por semana e em pessoas que faziam exercícios com frequência menor, comparando estatisticamente os riscos por meio do teste do qui-quadrado.

Teste *t* (ou teste *t* de Student). Teste estatístico utilizado para determinar se o valor médio de uma variável contínua em um grupo difere significativamente daquele em outro grupo. Por exemplo, entre participantes de um estudo que foram tratados com dois antidepressivos diferentes, um teste t poderia ser utilizado para comparar os escores médios de depressão após o tratamento nos dois grupos (teste *t* para duas amostras não pareadas) ou a mudança média a partir da linha de base até após o tratamento

nos dois grupos (teste *t* para duas amostras pareadas). Ver também *teste t para uma amostra* e *teste t para duas amostras*.

Teste *t* para duas amostras. Teste estatístico usado para comparar o valor médio de uma variável em uma amostra com seu valor médio em outra amostra. Por exemplo, investigadores descobriram que participantes tratados com suplementos de azeite de oliva tiveram um aumento médio de 10 mg/dL nos níveis de HDL durante o estudo quando comparados com um aumento de 2 mg/dL naqueles tratados com placebo ($P = 0{,}14$, utilizando o teste t para duas amostras). Ver também *teste t para uma amostra*.

Teste *t* para uma amostra. Teste estatístico utilizado para comparar o valor médio de uma variável em uma amostra com uma constante fixa (um determinado número). O tipo mais comum de teste *t* para uma amostra é um *teste* t *pareado*, no qual a média da amostra para a diferença entre medições pareadas (p. ex., no mesmo sujeito em momentos diferentes no tempo) é comparada com zero. Por exemplo, investigadores descobriram que médicos homens ganharam uma média (± DP) de 4 ± 3 kg de peso durante o período de sua residência ($P = 0{,}03$, utilizando um teste *t* para uma amostra). Ver também teste *t* para duas amostras.

Teste *Z*. Teste estatístico usado para comparar proporções de modo a determinar se elas são diferentes uma da outra de forma estatisticamente significativa. Ao contrário do teste do qui-quadrado, que é sempre bilateral, o teste *Z* pode ser usado para hipóteses unilaterais. Por exemplo, um teste *Z* unilateral pode ser usado para determinar se a proporção de presidiários com diabetes é significativamente maior que a proporção de pessoas não reclusas que têm diabetes. Da mesma forma, um teste *Z* bilateral (ou um teste do qui-quadrado) poderia ser usado para determinar se a proporção de presidiários com diabetes é significativamente diferente (i.é., menor ou maior) do que a proporção de pessoas não reclusas com diabetes.

Validade aparente. Termo que descreve a capacidade de uma aferição em medir um determinado fenômeno, com base na plausibilidade. Em geral, não é um método muito confiável para avaliar a validade. Por exemplo, uma medida que avaliava se um adolescente era popular parecia ter validade aparente, pois os investigadores achavam que ela diferenciava os alunos populares no ensino médio daqueles que não eram populares. Ver também *validade de construto*, *validade de conteúdo* e *validade de critério*.

Validade de construto. Um termo que descreve quão bem uma aferição corresponde às definições teóricas do traço (“construto”) que está sendo medido. Por exemplo, acreditava-se que uma medida de ansiedade social apresentava validade de construto porque havia diferenças substanciais nos seus valores entre pessoas cujos amigos descreviam a si próprios como “gostando de diversão” e “extrovertidos” quando comparados com aqueles que se descreviam como “tímidos” ou “com pouca probabilidade de frequentar festas”. Ver também *validade de conteúdo* e *validade de critério*.

Validade de conteúdo. Um termo que descreve quão bem uma medida representa diversos aspectos do fenômeno em estudo. Por exemplo, acreditava-se que uma medida da insônia tinha validade de conteúdo, pois ela mediu a quantidade total de sono, os episódios de despertar noturno, os episódios de despertar cedo na manhã, o nível de energia ao acordar e a sonolência diurna. Ver também *validade de construto* e *validade de critério*.

Validade de critério. Termo que descreve quão bem uma medida se correlaciona com outras formas de medir o mesmo fenômeno. Por exemplo, uma medida de depressão em adolescentes tinha validade de critério, pois se correlacionava fortemente com escores do inventário de depressão de Beck. Ver também *validade de construto* e *validade de conteúdo*.

Validade preditiva. Termo que descreve quão bem uma medida representa o fenômeno que pretende medir, com base na sua capacidade de predizer desfechos relacionados. Por exemplo, a validade

preditiva de uma medida da depressão seria fortalecida se ela estivesse associada com o risco subsequente de suicídio.

Validade. Grau em que uma medida representa o fenômeno de interesse. Por exemplo, o escore em um questionário sobre qualidade de vida é válido na medida em que ele realmente mede a qualidade de vida.

Valor *P*. Baseado em testes estatísticos, é a probabilidade de encontrar, tão somente pelo acaso, um efeito (ou, mais precisamente, um valor da estatística do teste) de tamanho igual ou superior ao encontrado no estudo se a hipótese nula for realmente verdadeira. Por exemplo, se a hipótese nula for que tomar café não está associado com o risco de infarto do miocárdio, e o estudo tiver mostrado que o risco relativo de infarto do miocárdio entre pessoas que tomam café em comparação com pessoas que não tomam café é de 2,0 com um valor P de 0,10, houve 10% de probabilidade de encontrar um risco relativo de 2,0 ou mais nesse estudo se não houvesse associação entre café e infarto na população.

Valor preditivo negativo. Probabilidade de que uma pessoa com um resultado de teste negativo não tenha a doença que está sendo testada. Por exemplo, em uma população de homens com uma prevalência de câncer de próstata de 10%, o valor preditivo positivo de um antígeno prostático específico (PSA) > 4,0 ng/mL é de aproximadamente 30%. Ver *prevalência*, *sensibilidade* e *especificidade*.

Valor preditivo positivo. Probabilidade de que uma pessoa com um resultado de teste positivo tenha a doença que está sendo testada. Por exemplo, em uma população de homens com prevalência de câncer de próstata de 10%, o valor preditivo positivo de um antígeno prostático específico (PSA) > 4,0 ng/mL é de aproximadamente 30%. Ver *sensibilidade* e *especificidade*.

Valores marginais. Os valores totais das linhas e colunas em uma tabela de contingência. Por exemplo, examinando os valores marginais na tabela 2 x 2, foi possível constatar que havia números similares de homens e mulheres no estudo.

Variabilidade. Grau de dispersão em uma medida, geralmente calculado como o desvio padrão. Por exemplo, se a mudança no peso corporal produzida por uma dieta varia de ganho substancial de peso a perda substancial de peso, a mudança é muito variável. Ver também *desvio padrão* e *erro padrão da média*.

Variáveis categóricas policotômicas. Variáveis categóricas com três ou mais categorias. Por exemplo, o tipo sanguíneo, que inclui A, β e O, é uma variável categórica policotômica.

Variável categórica. Variável que pode ter apenas um número limitado de valores possíveis. Por exemplo, o investigador transformou suas aferições de escolaridade autorrelatada em uma variável categórica com quatro valores: menos do que ensino médio completo, ensino médio completo ou superior incompleto, ensino superior completo, pós-graduação. Ver também *variável contínua*, *variável dicotômica* e *variável ordinal*.

Variável confundidora ou de confusão. Ver *confundimento*.

Variável contínua. Uma medida que, teoricamente, pode ter um número infinito de valores possíveis. Na prática, o termo é, com frequência, utilizado para medidas que têm “muitos” valores possíveis (alguns dizem 10 ou mais, outros dizem 20 ou mais). Por exemplo, a pressão arterial sistólica foi aferida como variável contínua em mmHg utilizando um esfigmomanômetro com coluna de mercúrio. Ver também *variável categórica*, *variável dicotômica* e *variável discreta*.

Variável de desfecho. Definição formal do desfecho para cada sujeito. Por exemplo, em um estudo sobre os efeitos de diferentes tipos de exercícios físicos sobre o peso e a composição corporal, as variáveis de desfecho foram definidas como a mudança no peso em kg desde a linha de base até a

aferição final após um ano e a mudança na circunferência da cintura em cm durante o mesmo período.

Variável dependente. Ver *variável de desfecho*.

Variável dicotômica. Variável que pode ter um de dois valores possíveis, como sim/não ou masculino/feminino. Por exemplo, o examinador dicotomizou a pressão arterial sistólica em hipertensiva (≥ 140 mmHg) ou não. Ver também *variável categórica* e *variável contínua*.

Variável discreta. Tipo de variável que aceita apenas valores inteiros. Para fins práticos, as variáveis contínuas são às vezes tratadas como discretas. Por exemplo, a idade geralmente é expressa como idade em anos no último aniversário, e o consumo atual de tabaco como a média de número de cigarros fumados por dia. Ver também *variável contínua*.

Variável independente. Ver *variável preditora*.

Variável instrumental. Variável associada ao preditor, mas não associada de outra forma ao desfecho; pode, portanto, ser usada para estimar indiretamente o efeito do preditor sobre o desfecho. Por exemplo, pesquisadores encontraram grandes diferenças regionais no uso de uma nova vacina contra influenza e, portanto, usaram o local da residência como variável instrumental para estudar o efeito da vacina contra influenza na mortalidade total em idosos.

Variável nominal. Variável categórica para a qual não há ordem lógica. Por exemplo, a afiliação religiosa (Cristã, Budista, Hindu, Muçulmana, Judaica, outra, nenhuma) foi codificada como variável nominal.

Variável ordinal. Variável categórica cujos valores têm uma ordem lógica. Por exemplo, o consumo atual de álcool foi tratado em um estudo como variável ordinal. Os valores eram: nenhum consumo de álcool, 1 ou 2 drinks por semana, > 2 mas < 7 drinks por semana, 1 a 2 drinks por dia e ≥ 3 drinks por dia. Ver também *variável nominal*.

Variável preditora. Ao considerar a associação entre duas variáveis, é aquela variável que ocorre primeiro ou que, pelo conhecimento da biologia da associação, tem maior probabilidade de causar a outra. Por exemplo, em um estudo para determinar se a obesidade está associada a um risco aumentado de apneia do sono, a obesidade seria a variável preditora. Em um ensaio clínico randomizado analisado por meio da análise de intenção de tratar, a variável preditora é o grupo designado.

Variável. Medida que pode ter diferentes valores. Por exemplo, o sexo é uma variável, pois pode ter dois valores diferentes – masculino e feminino. Ver também *variável categórica*, *variável confundidora*, *variável contínua*, *variável dicotômica*, *variável discreta*, *variável nominal*, *variável ordinal*, *variável de desfecho* e *variável preditora*.

Viés. Erro sistemático em uma aferição, ou na estimativa de uma associação, devido a uma falha no delineamento, na execução ou na análise de um estudo. Por exemplo, devido a um viés na forma como os sujeitos se lembravam de sua exposição a substâncias tóxicas, pacientes com leucemia tinham maior probabilidade de relatar uso prévio de inseticidas do que os controles.

Viés de amostragem. Erro sistemático que faz com que a amostra de pessoas incluídas no estudo não represente a população-alvo. Por exemplo, se os participantes de um estudo sobre fatores de risco para osteoporose fossem recrutados entre pacientes hospitalizados por fratura de quadril, a queda poderia aparecer falsamente como um fator de risco para osteoporose em razão do viés de amostragem.

Viés de duplo padrão-ouro. Ver *viés de verificação diferencial*.

Viés de espectro. Situação na qual a acurácia de um teste é diferente na amostra do que teria sido na população porque o espectro da doença (que afeta a sensibilidade) ou da não doença (que afeta a especificidade) é diferente daquele encontrado na população onde o teste será usado. Por exemplo,

devido ao viés de espectro, um novo teste sérico delineado para diagnosticar câncer de esôfago apresentou acurácia relativamente elevada em um estudo com pacientes com câncer de esôfago avançado comparados com estudantes de medicina saudáveis, mas teve desempenho ruim quando utilizado em pacientes idosos com disfagia sem diagnóstico definido.

Viés de não resposta. Tipo de viés no qual deixar de responder (p. ex., não responder um questionário) afeta os resultados de um estudo. Por exemplo, os investigadores estavam preocupados com o viés de não resposta no seu estudo sobre o uso de drogas ilícitas sobre o risco de desenvolver insuficiência renal.

Viés de publicação. Distorção da literatura publicada que ocorre quando os estudos publicados não são representativos da totalidade de estudos que foram realizados, geralmente porque os resultados positivos são submetidos e publicados com maior frequência do que os resultados negativos. Por exemplo, os autores de uma metanálise suspeitaram de viés de publicação quando descobriram que seis estudos positivos pequenos, mas apenas um estudo negativo de grande porte, haviam sido publicados.

Viés de verificação diferencial. Viés que ocorre em estudos sobre testes diagnósticos quando diferentes padrões-ouro são aplicados a sujeitos diferentes, dependendo em parte do resultado do teste que está sendo estudado. Por exemplo, em um estudo sobre o rastreamento com antígeno prostático específico (PSA) em homens, aqueles com níveis mais elevados de PSA foram submetidos a biópsias prostáticas, e aqueles com níveis normais foram seguidos clinicamente. Isso levantou a preocupação de que o viés de verificação diferencial poderia ter aumentado falsamente a sensibilidade e reduzido a especificidade do rastreamento com PSA em homens com câncer de próstata indolente.

Viés de verificação. (Também denominado viés de investigação diagnóstica ou viés de referência). Viés na avaliação da acurácia de um teste que ocorre quando os participantes realizam de forma seletiva a verificação por um padrão-ouro com base em parte nos resultados do próprio estudo. Por exemplo, se um estudo sobre a acurácia da percussão torácica para diagnosticar pneumonia incluísse apenas pacientes que tinham realizado uma radiografia de tórax, e se aqueles com macicez na percussão tivessem maior probabilidade de realizarem uma radiografia, a sensibilidade da percussão seria falsamente elevada, e a especificidade, falsamente reduzida, devido ao viés de verificação.

Viés diferencial. Termo genérico para a situação na qual uma medida varia sistematicamente de acordo com o estado do sujeito, em geral se ele é caso ou controle; ocorre com maior frequência com exposições que precisam ser lembradas. Por exemplo, uma vez que os casos adultos de doença celíaca tinham maior probabilidade de se recordarem de exposições na infância a produtos contendo trigo do que seus irmãos que haviam crescido no mesmo domicílio, os investigadores suspeitaram de viés recordatório diferencial. Ver também *viés não diferencial*.

Viés do observador. Situação na qual um investigador (ou assistente de pesquisa) realiza uma avaliação não objetiva que é afetada por seu conhecimento de um ou mais dos atributos do sujeito, como se o sujeito é caso ou controle ou se foi exposto ou não a um determinado fator de risco. Por exemplo, o viés do observador foi aparentemente responsável pelo achado de que, com base em uma entrevista, os adolescentes hispânicos tinham maior probabilidade de serem caracterizados como tendo problemas de controle da irritabilidade do que os asiáticos, pois um inquérito autoadministrado e uma revisão dos registros escolares não encontraram diferenças entre os dois grupos.

Viés do sujeito. Ver *viés recordatório*.

Viés não diferencial. Tipo de viés que não é afetado pelo fato de o sujeito ser caso ou controle (ou pelo fato de um sujeito ter sido exposto ou não a uma terceira variável). O viés não diferencial tende a tornar as associações mais difíceis de serem encontradas, pois reduz as diferenças aparentes entre os grupos. Por exemplo, embora recordar-se de exposições prévias a antibióticos era imperfeito tanto em casos

quanto em controles, o viés pareceu ser não diferencial, pois uma revisão de registros médicos mostrou que ambos os grupos tinham inacurácias similares. Ver também *viés diferencial*.

Viés recordatório. Tipo específico de viés no qual o fato de um sujeito lembrar ou não de um fator de risco que ocorreu no passado e a forma como ele se lembra são influenciados por outro fator, especialmente o fato de o sujeito ser caso ou controle. Por exemplo, acreditou-se que o viés recordatório era o motivo pelo qual os casos de esclerose lateral amiotrófica tinham maior probabilidade de lembrarem de exposição a inseticidas do que os controles.

---

[1]N. de R.T. SF, do inglês short form, significa forma curta

Conheça também:

FLETCHER, R.W.; FLETCHER, S.E.; FLETCHER, G.S. – Epidemiologia clínica. 5.ed.
GLANTZ, S.A. – Princípios de bioestatística. 7.ed.
GREENHALGH, T. – Como ler artigos científicos: fundamentos da medicina baseada em evidências. 4.ed.
GUYATT, G.; RENNIE, D.; MEADE, M.O.; COOK, D.J. – Diretrizes para utilização da literatura médica: fundamentos para prática da medicina baseada em evidências. 2.ed.
GUYATT, G.; RENNIE, D.; MEADE, M.O.; COOK, D.J. – Diretrizes para utilização da literatura médica: manual para prática da medicina baseada em evidências. 2.ed.
HULLEY, S.B.; CUMMINGS, S.R.; BROWNER, W.S.; GRADY, D.G.; NEWMAN, T.B. – Delineando a pesquisa clínica. 4.ed
LOPES, R.D.; HARRINGTON, R.A. – Compreendendo a pesquisa clínica
PERERA, R.; HENEGHAN, C.; BADENOCH, D. – Ferramentas estatísticas no contexto clínico
POPE, C.; MAYS, N. – Pesquisa qualitativa na atenção à saúde. 3.ed.
ROTHMAN, K.; GREENLAND, S.; LASH, T. & COLS. – Epidemiologia moderna. 3.ed.
SAMPIERI, R.H.; COLLADO, C.F.; LUCIO, M.P.B. – Metodologia de pesquisa

---

grupo a
EDUCAÇÃO
Mc Graw Hill Education
bookman
artmed
penso
artes médicas

DELINEANDO A PESQUISA CLÍNICA

4ª Edição

STEPHEN B. HULLEY
STEVEN R. CUMMINGS
WARREN S. BROWNER
DEBORAH G. GRADY
THOMAS B. NEWMAN